# THESAURUS AS A TOOL FOR THE MANAGEMENT OF GOVERNMENT INFORMATION

A Paper Prepared by the
National Archives of Canada for the

Thesaurus Standards Working Group
Treasury Board Information Technology
Standards Programme (TBITS)

October 1996

# PREFACE

This document is a consolidation of the suite of reports in consultation draft which were tabled at the meeting of the TBITS Thesaurus Standards Working Group on June 14, 1994. The suite of reports included the following: *Definition of Requirements for Thesauri in Government; Thesaurus: A Guide to the Use of Thesaurus in the Management of Government; Information Holdings, Inventory of Thesauri, Thesaurus Projects and Controlled Vocabularies in the Federal Government; Thesaurus Standards: an Annotated Checklist; and Thesaurus: an Annotated Bibliography.* On the recommendation of the Working Group at its fall meeting in 1994, the reports have been consolidated into one document.

The document has been revised to reflect comments received from members of the Working Group as a whole. As well at the fall meeting three members (Editorial Sub-committee, Appendix A) offered to assist with the revision of the reports. Two meetings were held. At the most recent meeting of the Editorial Sub-committee, held in July 1996, the need for functions access, as distinct from subject access, was questioned; as well, the validity of the use of thesaurus to provide functions access was queried. These are important issues which require reflection, discussion, debate.

**By the functions approach is meant organisation and access of records and information by business functions, programmes, and activities. At a recent seminar on Canadian privacy legislation, it was stated that members of the Canadian public increasingly require access to Government of Canada information by function and programme. If this is so, it follows that the functions approach would serve to facilitate public access, as well as to facilitate the effective management of government records and information. By functions thesaurus is meant a thesaurus of functions terms. The National Archives of Australia, has, for a number of years, used a functions thesaurus. The concept of functions thesaurus, in itself, is thus a proven one. Therefore, it would seem that discussion on these issues should focus, not so much on the relative merits of the two approaches (functions and subject), but rather on where and how each approach may best apply.**

This document, which is in distribution to members of the Thesaurus Standards Working Group, will encourage thought and discussion of the above ideas. Members may carry forward the discussion of these concepts to the information management community at large. If so, it will have served the purpose of heightening awareness about issues important to the information society and to the construction of the information highway in Canada.

# THESAURUS AS A TOOL IN THE MANAGEMENT OF GOVERNMENT INFORMATION

## TABLE OF CONTENTS

PREFACE

1. INTRODUCTION . . . 5
    Background . . . 5
    About this Document . . . 6

2. WHY BUILD A THESAURUS IN THE FEDERAL GOVERNMENT . . . 8
    The Objectives of Constructing a Common Thesaurus Government-Wide . . 10
    Government Information Locator Service . . . 10
    The Vision - a Government Information Locator Service . . . 11

3. PROPOSED GENERIC REQUIREMENTS FOR THESAURI IN GOVERNMENT . . . 12
    Multilingual Capacity . . . 12
    Language Equivalencies - Linguistic and Legal . . . 12
    Structure . . . 12
    Conceptual Framework . . . 12
    Technology . . . 13
    Flexibility . . . 13
    Use . . . 13
    Training, User Support, and Documentation . . . 14
    Management . . . 14
    Informatics Requirements Example . . . 15
    Functional Requirements . . . 15

4. THESAURUS IN THE IM LANDSCAPE . . . 16
    Corporate Memory and Thesaurus . . . 18
    Legislation . . . 19
    Policy . . . 19
    Management of Government Information Holdings . . . 20
    Identifying Information Management Requirements . . . 20

5. FUNCTIONS THESAURUS - WHAT IS IT? . . . 24
    Functions Thesaurus . . . 24

6. **DEFINITION OF CONCEPTS** . . . 29
  - Thesaurus . . . 29
  - Thesaurus Characteristics . . . 29
  - Controlled Vocabulary . . . 30
  - Thesaural Approach . . . 31
  - Thesaurus Structure . . . 31
  - Thesaurus Display . . . 31
  - Thesaurus Construction/Design . . . 32
  - Multilingual Thesaurus . . . 33
  - Retrieval Aspects . . . 33
  - Term Selection . . . 34
  - Thesaurus Maintenance . . . 34
  - User Interface . . . 34
  - Training . . . 34
  - Automation . . . 35
  - Faceted Thesaurus . . . 35
  - Common Thesaurus or Metathesaurus . . . 36

7. **CHECKLIST OF FACTORS TO CONSIDER IN ASSESSING THESAURI** . . 41

8. **INVENTORY OF THESAURI, THESAURUS PROJECTS AND CONTROLLED VOCABULARIES IN THE FEDERAL GOVERNMENT** 46

9. **STANDARDS FOR THESAURUS, TERMINOLOGY, VOCABULARY AND INDEXING** . . . 122

10. **THESAURUS: AN ANNOTATED BIBLIOGRAPHY** . . . 129

**Appendix A** TBITS Thesaurus Standards Working Group: Official Documents

# 1. INTRODUCTION

## Background

The Treasury Board Information and Technology Standards (TBITS) Thesaurus Standards Working Group is a joint initiative of the National Archives of Canada and the Treasury Board of Canada which developed out of projects and meetings about thesaurus from 1991 to 1993. For a number of years the idea of investigating the use of a thesaurus to manage government information had been discussed informally in the information management community. As part of a partnership initiative called IMOSA[1] (established by the National Archives of Canada and the Department of Communications), a study[2] was commissioned to investigate the feasibility of common thesaurus in the Government of Canada. The study was conducted by a research team from Laval University. Results indicated that there was considerable interest in the Government of Canada in the development of a common thesaurus and it was considered to be feasible to develop a common thesaurus. Research reported at the American Society of Information Science also indicated a growing interest in the concept of common and meta thesaurus.[3]

In June 1992, a meeting was organised by the Information Management Standards and Practices Division of the National Archives of Canada to discuss the need for thesauri. In October, 1992 a second meeting was held to discuss thesaurus issues. It was decided that there was sufficient need and interest to organise an interdepartmental Thesaurus Working Group under the Treasury Board Information and Technology Standards (TBITS) programme.

The National Archives of Canada, which is responsible for facilitating the management of government records and for assisting the Treasury Board in the implementation of the Management of Government Information Holdings Policy (MGIH), was asked to chair the TBITS Thesaurus Standards Working Group on behalf of the TBITS programme. The purpose of the working group is to explore the use of thesaurus as a tool to manage government information.

A call letter was sent out by Treasury Board in September 1993. A first meeting was held in October 1993 and the terms of reference and work plan were approved. A second meeting was held in November at which time three sub-groups were established: Group I: Definition of

---

[1] IMOSA: Information Management and Office Systems Advancement. A joint initiative of the National Archives of Canada Government Records Branch and the Department of Communications Canadian Workplace Automation Research Centre (now called Centre for Information Technology Innovation). Overview document. Revised. Nov., 1991.

[2] Rapport de l'Étude de faisabilité: Développement d'un modèle pour un thésaurus de base informatisé. Un projet conjoint des Archives nationales du Canada, Direction des documents gouvernementaux et du Ministère des Communications, Centre de recherche sur l'informatisation du travail, novembre 1992. 158 p.

[3] ASIS '93. Proceedings of the 56th ASIS Annual Meeting, 1993 Volume 30, Columbus, Ohio, October 24-28, 1993. Published for the American Society for Information Science by Learned Information, Inc., Medford, New Jersey, 1993. 334 p.

Requirements, Group II: Thesaurus Standards Review, and Group III: Controlled Vocabulary and Thesaural Approach. Members self-selected into three sub-groups. Each group was responsible for a part of the work plan established at the first meeting of the TBITS Thesaurus Standards Working Group (TSWG). Terms of reference and membership of the Working Group are given in Appendix A.

With the support of National Archives staff, the three sub-groups established a number of projets to investigate concepts associated with thesauri and their existing applications in government. The objectives were: to outline how a common thesaurus could be used in government and how it relates to information management; to define requirements for thesauri in government; to develop a thesaurus standards annotated checklist, a thesaurus annotated bibliography as well as a list of factors to consider in assessing thesauri, and, finally, to identify existing thesauri and controlled vocabularies in use in the Government of Canada.

**About this Document**

This document was prepared in 1993-94 with the help of staff of the National Archives of Canada. The paper includes:

1. A statement of generic requirements for thesauri in government (Chapter 2);
2. An outline on how a thesaurus can be used in government and how its use relates the information management environment (Chapters 3, 4 and 5);
3. A list of factors to consider in assessing thesauri (Chapter 7);
4. An inventory of thesaurus, thesaurus projects, and controlled vocabularies in the Government of Canada (Chapter 8);
5. A checklist of standards (Chapter 9);
6. A bibliography on thesaurus (Chapter 10) and;
7. A definition of concepts (Chapter 6).

Members of the TSWG Sub-group I: Definition of Requirements discussed how thesauri and thesaural approaches (use of controlled vocabularies) can serve as tools to manage government information. The idea of a common, government-wide, thesaurus as a major building block in a **government information locator service** was developed. See Chapter 4: Why Build a Thesaurus in the Federal Government. The concept of a functions thesaurus was introduced. It was suggested that a functions thesaurus, that is a thesaurus constructed using business functions, could be a key element in a government-wide thesaurus. See Chapter 3: What Is a Function Thesaurus for a discussion of this concept. These concepts were brought forward to the Working Group as a whole when the reports were tabled on June 4, 1994. See the preface for further information in the development and revision of this report.

This document distributed to the TSWG may serve to initiate debate on the various approaches to managing government records and information. It could also serve as a reference tool and a guide for information management professionals concerned with the identification of solutions for managing government information and who wish to familiarize themselves with the concept of thesaurus. The Information Management Standards and Practices Division of the National Archives of Canada has used this information in providing advice concerning the description of records. As well, the document promotes the sharing of information by identifying existing applications of thesauri and controlled vocabularies in government. See Chapter 8: Inventory of Thesauri, Thesaurus Projects and Controlled Vocabularies in the Federal Government. The report may also promote the building of common systems by suggesting a set of generic requirements for future thesaural endeavours in government. See Chapter 2.

Comments and questions regarding this report may be addressed to the Thesaurus Standard Working Group Chair:

Rosemary Murray-Lachapelle
National Archives of Canada
395 Wellington Street
Ottawa, Ontario
K1A 0N3
Telephone: (613) 947-1512
Facsimile: (613) 947-1500
Internet: rmurray-lachapel@archives.ca

## 2. WHY BUILD A THESAURUS IN THE FEDERAL GOVERNMENT

(Section prepared by TSWG Sub-group I: Definition of Requirements)

Why should the federal government have an interest in building a thesaurus? The goal is to provide intelligent access and ease of use government-wide. Decision makers are responsible for programme delivery and are held accountable for resources used and services rendered. In order to be productive and accountable, they require information. Information is a major national resource as has recently been stated in the Blueprint document. "Government information in all forms (e.g. print, voice, electronic, or image), is a strategic resource and will be effectively managed throughout its lifecycle.... It will be necessary to implement mechanisms to easily and accurately find government information. Benefits include improved availability and quality of information for processing and decision making, resulting in improved service."[4] In order for managers to have access to the information they require, the right information must be planned for, created, identified, described, and indexed in such a way as to be available, understandable, and useable over time. This in turn presupposes quality control: information must be accurate, relevant, current; it must by characterised by integrity and be situated in context. A thesaurus facilitates the management of concepts. By providing the right information at the right time, it can be used by managers as a strategic tool to support the decision-making process, to assist in the satisfaction of operational business needs, in the fulfilment of accountability requirements, and ultimately contribute to the preservation of the corporate memory.

Retrieval of information using a thesaurus increases the relevance and precision of search results (see Appendix D). In the future it is likely that such automated search techniques as full and free text with intuitive search capabilities will operate in conjunction with thesaural applications to provide more effective organisation and retrieval of information and information holdings. This type of product is already emerging on the market.

The Australian Public Service has put the matter succinctly in its publication **Management of Electronic Documents**: "...retrieval software now exists that links terms from a controlled thesaurus to a text document by the determination of what concepts are contained in the document, as built on evidence from the full text of the document. This type of software accumulates evidence and provides a graded result, with the documents presented in order of importance. For a searcher, the advantage is in the existence of the thesaurus. Thesauri commonly provide a list of preferred terms, as well as a hierarchical structure, so that relationships between terms are provided (e.g. broader and narrower terms). The searcher has available the vocabulary used either by the human indexer or the retrieval system, and therefore has some chance of common understanding of the words used in the indexing.

---

[4] Blueprint for Renewing Government Services Using Information Technology: Discussion Draft. Ottawa: Treasury Board of Canada, Secretariat. 1994. p. 10.

From the point of view of management of corporate information, the use of a well-constructed thesaurus for indexing that reflects how words are used within an agency enhances the common understanding of terms and thus contributes to efficient intra-agency communication. It follows that the use of the same thesaurus for the indexing of all information in an agency will establish logical and intellectual links between information kept in different media, e.g. library catalogue, paper files in records management and on-line databases for data and documents. The construction of thesauri and their styles and uses will not be explored at this point. Suffice it to say that there is a rich body of research literature that can be called upon as well as specialists in the field for assistance. Keywords should be used in electronic document management with the important proviso that an agency has corporately managed standards, including a thesaurus."[5]

The ultimate objective is to effectively describe information in order to make it available, useable, and understandable over time, not to lose the most important identifying tags in the information, and to do so in a cost-effective manner. The thesaural approach including the thesaurus structure per se, when broadened beyond subject concepts, to include corporate names, for example, enables the creator of information to capture or tag at the point of creation, the data essential to the identification, description, and classification of information for purposes of fulfilling business functions, to respond to accountability requirements, and to preserve the corporate memory. Ultimately, "subject" is an interpretation. Known facts are not. Therefore the facts representing information on who, where (sector of the department,) representing what function, what mandate, what laws, as well as the why, must be linked to each other and to all other related information.

Finally, it is clear that a variety of benefits would adhere to the creation and use of a common government-wide thesaurus. For example government would invent once and use many times. Information would be identified as existing perhaps for the first time. Managers would discover in some instances what is available through information identification, description, organisation, and sharing. Access to information would be improved and the mandate to disseminate information, and to share among disciplines would be realised. Savings through cost-sharing would be achieved. Decision making and programme delivery through improved information availability and useability would be enhanced. Risk reduction would be realised and information harnessed to support business needs.

The basic principle of the common thesaurus is simple: information is shareable - invent once; use many times. The **Blueprint** summarises the benefits as follows:

- Improved availability of information;
- Improved decision-making by program managers and policy-making as a whole, at both the strategic and operational levels;

---

[5] Management of Electronic Documents in the Australian Public Service. A Report prepared by the IESC's Electronic Data Management Sub-committee. Information Exchange Steering Committee (IESC). April, 1993. p. 25.

- Enhanced client service; and,
- Easier and speedier service delivery to all regions of the country.[6]

### The Objectives of Constructing a Common Thesaurus Government-Wide

The objective of constructing a thesaurus is to retrieve relevant information in a cost effective manner in order to satisfy business needs and to reduce the risks inherent in not having access to accurate information in a timely manner. The objective of constructing a government-wide or common thesaurus is to share information, to create once and use many times. A common thesaurus would be a major building block in the government information architecture which would enable decision-makers to reach information they need but are not at present able to ascertain is there; to identify the existence of the information by linking and mapping to information sources throughout government by building bridges among a wide variety of modules, holdings, or banks of information; to ensure consistency and quality of information; to facilitate the use of contextual information on the creation of records throughout the life cycle of records, including their description and disposition; to facilitate the exchange of information internally and externally; and to disseminate information.

### Government Information Locator Service

The working group thought of a **Government-wide Information Locator Service (GILS)** as a finding aid in which a thesaurus would be key. Neither a repository nor a huge central system, it would be a mapping device to government information resources. The objective of such a system is to provide access to the information required to fulfil business functions, to actually fill certain business functions by producing and packaging the information required, and to provide public access to government information which is, in itself, a business function of the Public Service of Canada. In the U.S. the National Archives and Records Administration, the General Services Administration, and the Office of Management and Budget are investigating such a model.[7] **Graphic I** provides an overview of this model as it could apply within the Government of Canada. At the top of Graphic I is situated the government-wide locator service which interfaces with the Canadian public and with the body of information generated and housed by the Government of Canada. The goal is **intelligent access** and **ease of use.** The government-wide system maps to departmental information locator services **(DILS)**. Major building blocks in the system are thesauri, common thesauri, and controlled vocabularies.

---

[6] Blueprint, op. cit., p.28.

[7] Using Z39.50 in an application for the Government Information Locator Service (GILS) January 24, 1994 Draft. Available on the Fedworld electronic bulletin board (703-321-8020 or by anonymous FTB (File Transfer Protocol) via the Internet at 130.11.48.10 as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCIItext format). The authoritative document describing the objectives, characteristics, et. of GILS is Government Information Locator Service. Eliot Christian, United States Geological Survey.

## The Vision - a Government Information Locator Service

The idea of using a thesaurus as a tool to **manage** information constitutes an expansion of the original concept of thesaurus. The original concept was to provide subject access to information. The newer concept is to use the thesaural structure for identification and description of information and to build a common thesaurus designed to act as a mapping tool linking to, and connecting with, various information banks such other thesauri, common thesauri (within departments for example), indexes, data dictionaries, data banks, etc. Within a department, for example, the thesaurus could be constructed from a subject/concept index, derived from a departmental classification scheme, which in turn would be derived from the departmental Operational Planning Framework (OPF). This developmental path would ensure the essential linkage to the functions and activities (business functions) of the department. The Functions Thesaurus of the Australian Archives is an example of a common thesaurus.[8] An example of the thesaural approach structure exists in prototype at Environment Canada[9] where a consolidated index vocabulary prototype has been developed to map terms among thesaural applications in records management operations, the library system, and Envirosource. Other applications exist today in various stages of development within the Government of Canada and elsewhere. Examples of them are listed in the **Inventory of Thesauri, Thesaurus Projects, and Controlled Vocabularies** in Appendix B. A government-wide locator service would map among the various departmental thesaural applications as well as among other information stores and would hence provide an important contribution to an information architecture for government.

---

[8] Functions Thesaurus, op. cit.

[9] Information on the Environment Canada initiatives are available in the following reports: Envirosource (TM) Intelligent Interface to Information Holdings, Thesaurus and Consolidated Index, Final Report, March 1993. Prepared by Consulting and Audit Canada for Environment Canada and Prototype Assessment: Thesaurus/Consolidated Index (Enivronment Canada), August 1993. Prepared by Information Ecosystems Group Inc. for Environment Canada.

## 3. PROPOSED GENERIC REQUIREMENTS FOR THESAURI IN GOVERNMENT

This section, prepared by the TSWG Sub-group I: Definition of Requirements, describes generic requirements, conceptual and technological, for thesauri. As well, reference is made to additional documentation which will provide further details regarding data modelling and functional requirements.

### Multilingual Capacity

An essential requirement of thesaural applications at the government-wide level is to have the capacity to accept and operate in, and with, many languages. This is a requirement of law as regards French and English, the two official languages of Canada. However, developments in the area of self-government of the aboriginal peoples of Canada, as well as in the area of free trade which introduces hispanophone countries, indicate a requirement for the capacity to add multilingual structures at a later date as necessary.

### Language Equivalencies - Linguistic and Legal

The thesaurus would need to generate language equivalencies between English and French with the capacity to create additional linguistic equivalencies as necessary. The linguistic equivalencies would need to take into account legal linguistic aspects as well. As an example, the English word mortgage translates in ordinary language use as hypothèque in French. However the French civil law word hypothèque does not translate to the English common law word mortgage, but has another significance in law.

### Structure

All law and policy requirements are to be respected. In Section 4, under the sub-heading Corporate Memory and Thesaurus a list of the pertinent legislation and policy is provided.

### Conceptual Framework

From the theoretical perspective, a functions thesaurus within a department would be developed from the Operational Planning Framework (OPF) which identifies functions and activities. From the OPF would be generated the classification scheme within which documents could be identified, described, organised, and classified. A faceted thesaurus would be developed from the department's classification scheme. The developmental flow therefore would be from OPF to classification to thesaurus.

From the government-wide perspective the locator service would follow a similar developmental pattern. The functions thesaurus would be key to this development and would provide the originating structure or government-wide framework, based as it would be on legislative and central agency policy documents. In addition the mapping or bridging thesaural structure would link the government-wide common thesaurus to the departmental thesaurus which in turn within each department would map to various thesauri, controlled vocabulary, data banks, and other organised information holdings or menus.

## Technology

The thesaurus must be developed in a flexible modular way in order to provide technological readiness points so that departments at different technological levels of development might take advantage of the tool. An important requirement would be for the thesaurus to operate on a variety of commonly used technology platforms and be so constructed as to use as many off the shelf or already developed components as possible. Development of software should be minimal. Use and promotion of the software exchange would be encouraged. As well the system would be constructed with a view to integrating information from a variety of sources such as voice messaging systems (VMS), E-mail, and AI systems. Worth noting is the Australian experience. They state: "An important particular aspect of the electronic document should be separately considered at this stage, i.e. the **non-text document** (image, graphic, spreadsheet, voice-mail message, videotape). Full text retrieval in these cases is severely limited or impossible, given today's readily available technology. In the case of a corporate store where the content is largely non-text, the use of keywords combined with thesaurus terms is strongly recommended."[10]

## Flexibility

The thesaurus needs to be flexible in all aspects both logical and conceptual, constructed, as previously stated, in a modular way to provide readiness points so that departments at varying points of technological development could use the thesaurus. It would accommodate different kinds of information in different formats (that is, in multi-media forms) and would integrate information from different technologies. It could be implemented on a majority of commonly used platforms. In summary the thesaurus would be a dynamic, evolving structure that is able to be developed further. It would be transportable both logically and conceptually because it would be based on a standardised scheme and technical platform.

## Use

It is a requirement of the common thesaurus that it be characterised by **ease of use**. The mapping process would be transparent to the user. It would be clearly user friendly and amenable to the one stop shopping approach or the electronic window concept. Good mapping, plain language, and strong search capabilities would be essential. KWIC (key word in context) and KWOC (key word out of context) as well as permuted search and Boolean logic search capabilities would be desirable. Strong user support would be required including an automated help or online help module.

---

[10] Management of Electronic Documents, op. cit., p. 25.

## Training, User Support, and Documentation

A training, user support and documentation programme must be an integral part of the process to develop, implement and use a common thesaurus. Given that users at different levels of knowledge and skills would make different uses of the thesaurus, the training programme would need to accommodate different skill levels and uses. A form of electronic tutorial (CBT or diskette) should be developed to assist users in their acquiring of the knowledge and skills in an autonomous fashion. This tutorial should begin by outlining the objectives and benefits derived from learning how to use such a thesaurus so as to encourage users to acquire and develop these skills and to exploit the thesaurus to its maximum potential. The help function on the system should be a natural extension of the training program and ultimately help users become independent. In addition to the tutorial, brief orientation sessions should be made available for those users who learn more effectively through human exchange. A user support hot line should be made available at all times to answer inquiries from users and to minimize the waiting period for assistance in resolving problems.

**Documentation** is required for **system managers** responsible for installing the functional thesaurus on a given system. The system manager document would be of a technical nature to respond to the technologist needs. User documentation would also be required to navigate clients through the use of the thesaurus. This documentation would be in the form of a **user guide**. The user guide should reflect the particular needs of different user group. For instance, if the thesaurus is automated, the user documentation and help should be available on-line to facilitate access of updated information. The third type of documentation required involves a set of **procedures for the maintenance** of the thesaurus. It is imperative to provide such documentation for consistency and accuracy as there will be a number of individuals responsible for updating and maintaining the thesaurus over time. All documentation and service must be provided in English and French in accordance with the Official Languages Law.

## Management

Arrangements to direct and maintain the system would be an essential requirement. Certain functions of the common thesaurus could be delegated to certain departments especially in the area of custodianship of standardised terminology. Maintenance responsibilities would be assigned as well. A Board of Directors or Advisory Board of key partners would advise the administrator on key policies and guidelines. Roles and responsibilities would be defined and assigned as well. A system of maintenance for regular and easy updating would be required to include an administrator to act as an authority or custodian of government-wide terms as an Office of Primary Interest - linking with departments, perhaps with departmental integrators to maintain terms.

## Informatics Requirements Example

A statement or requirements for a bilingual thesaurus has been developed for use in the construction of a thesaurus for Info Source. The document shows data modelling based on ISO and ANSI thesaurus standards. Info Source, as stated earlier, is designed to meet the government policy requirements of access and privacy. It is one example of a common thesaurus application. The requirements document prepared for Info Source illustrates the baseline standards-compliant methodology for initiating and deploying a bilingual thesaurus project, a high-level architectural model compatible with this document's vision of government-wide thesaurus accessibility and interoperability, the requirements for a baseline standards-compliant data architecture for a generic bilingual thesaurus, and the baseline standards-compliant methodology for selection, classification, and indexing terms in a bilingual thesaurus.[11]

## Functional Requirements

A functional requirements report has been prepared in the context of the Envirosource Partnership Project 1993/94 which will be of interest to the readers of this report as may other related EC reports.[12]

---

[11] Info Source Specifications for a Bilingual Standards-Compliant Thesaurus. March, 1994. Prepared by Consulting and Audit Canada for the Treasury Board Secretariat. Ottawa, 1994.

[12] Consolidated Index and Indexing and Access Vocabulary. Thesaurus Functional Requirements Report. Second Draft, March 1994. Prepared by Julia Hudson. Ottawa, 1994 and Thesaurus of Envirosource TM Descriptors Environment Canada Information Holdings, January, 1994. Prepared by Susan Fitzmaurice. Available from the Environment Canada Library.

## 4. THESAURUS IN THE IM LANDSCAPE

(Prepared by TSWG Sub-group I: Definition of Requirements)

### Thesauri and other tools to Facilitate Intellectual Control of, and Intelligent Access to, Government Data, Information and Knowledge.

"Data are undifferentiated facts without context. Information is organized data that we, as individuals, have not yet absorbed. Once we have integrated information into our own internal frameworks, it becomes knowledge."[13] The business of government generates vast amounts of data, information and knowledge. A thesaurus is a tool which can facilitate intellectual control of and intelligent access to these vital assets.

To understand the unique role of a thesaurus, it is useful to situate it in relation to other information management tools in government. Perhaps the most widely known tool for maintaining intellectual control over government data is the **data dictionary.**

"A **data dictionary** is a facility for maintaining descriptions about the data held by an organisation. Typically it records the names, descriptions, size and structure of each data item. Data dictionaries were developed mainly to ensure standards in data naming conventions and to enable the impact of proposed changes to the database to be analysed... The information in a data dictionary is normally stored as a database itself."[14]

"Most government departments have several data dictionaries, each serving a particular function, program or an application. These dictionaries are usually neither rationalized nor linked together, often resulting in inconsistent definitions and redundant entries. The sources of information for definitions of data is (sic) not often documented in dictionaries... A few departments are beginning to give attention to the need for a departmental data dictionary. Some departments have recently initiated projects which would map their application portfolios against data dictionaries to create common data definitions."[15] "A standard set of definitions and values of... data elements is considered an essential building block of an infrastructure which supports the cross-functional information requirements of responsibility centre managers, program managers, departmental executives and central agencies."[16] "Without the availability of standard definitions and values of... data elements, it is difficult to develop open systems (i.e. portable and inter-operable systems) *Management of Common Reference Data Definitions*, undertake administrative re-engineering, achieve cross-functional integration, implement electronic data interchange applications and other electronic office functions, and create applications such as executive information systems, information holdings management, emergency preparedness. On a government-wide basis, it is, without having access to common definitions, difficult to develop and implement common administrative systems and provide multi-departmental common services to clients of the federal government."[17]

---

[13] Calamai, Peter, "Seeking Knowledge from a Flood of Information", The Ottawa Citizen; August 28, 1993, p. A8.

[14] Managing Information as a Resource. Central Computer and Telecommunications Agency, HMSO: London, 1990, p. 52.

[15] Comptroller General, Treasury Board of Canada. Ottawa, October 27, 1992, p. 5.

[16] Ibid, p. 1.

Another information management tool which facilitates both intellectual control of and intelligent access to government information is the **corporate inventory of information holdings.** The Management of Government Information Holdings (MGIH) policy requires institutions to maintain, "a current, comprehensive and structured identification or classification system or systems which provide an effective means for organizing and locating information and, in composite form comprise a corporate inventory for managing the institution's information holdings."[18]

"... a corporate inventory is a management tool that provides a comprehensive overview at the corporate level of how the information systems permit all the Information Holdings, regardless of their form or media, to be organised and managed in support of the programs and activities of the institution and allow government-wide legal and policy responsibilities to be met... it aids the institution in locating information for decision-making purposes, and in sharing information across its own program areas and across government. It is also fundamental to the application of information laws and policies."[19]

Ideally, a corporate inventory should relate individual holdings to the department's programs as they are described in its Operational Planning Framework (OPF) and provided to Parliament in Part III of the Estimates. This allows a department's information holdings to be understood in relation to the governmental programs and activities they support, adding meaning and providing context. Additional value derives from identifying holdings by type: common or government-wide, institution-wide, or program-specific.

In contrast to a corporate data dictionary which facilitates the management of data, and a corporate inventory which facilitates the management of information holdings, a **thesaurus** facilitates the management of concepts contained in a text. All thesauri are controlled vocabularies but not all controlled vocabularies are thesauri. What distinguishes a thesaurus from other controlled vocabularies, are the relationships provided among indexing terms. In a thesaurus, indexing terms are arranged hierarchically to show the relationships among them. The strength of the thesaural structure is in the relationships among concepts including the relationships of one to many and many to one which it permits. This means that the thesaurus can increase considerably the level of precision in information retrieval.

---

[17] Ibid, p. 1.

[18] Corporate Inventory of Information Holdings: Guide to the Structure, Data Model, Data Dictionary. Working Draft. Administrative Policy Branch, Information Management Practices, Treasury Board Secretariat. Ottawa, January 1991, p. 1.

[19] Ibid, p.1.

The strength of the thesaural structure is in the relationships among concepts (one to many and many to one) which it permits. A thesaurus based on functions, rather than on subject typology, relating to agencies and series[20] as in the Functions Thesaurus of the Australian Archives,[21] links offices horizontally and vertically through functional relationships, and provides an excellent tool for navigating through a government information system. The subject concepts in a functions thesaurus derive from the functions as defined in legislation and policy. As an example: Intelligence Agency: intelligence, counter-intelligence, espionage, spying, etc. linked to agencies and series with responsibilities in these areas. It is to be noted that thesauri based on subject typologies when used as indexing devices to periodicals and other documentation also have a significant place in the overall information landscape.

All these tools help facilitate intellectual control of, and/or intelligent access to, government data, information holdings and the concepts to which they relate. But is it possible to provide **intellectual control of and intelligent access to, the body of governmental knowledge?** That is to say, can government information, viewed within the context of its creation, use and significance over time, be made available at the desktop when and where it is needed? Can it be linked to the processes and human expertise which give it real meaning and dimension? Can value be added, retained and enhanced with each successive use? Will government decision-makers be able to draw upon the **corporate memory** of their organisation, tapping into not only data and information, but the body of knowledge, expertise, and experience of their organisations? Some departments are looking ahead to the day when technology will make possible such an integration of information and context. Government-wide common information management initiatives, are an important step in this direction. Thesauri are major building blocks along the road to the construction of common information management initiatives such as a government information locator service to provide access to information across government.

## Corporate Memory and Thesaurus

"In any organization, management of information is a vital part of day-to-day operations. The ability to retrieve, share and use information is critical not only for good internal communication within the government network, but ultimately it enables a federal government organisation to provide a high standard of service to the public."[22]

---

[20] The ICA Dictionary of Archival Terminology 2nd Ed. (1988), defines series as: Item or documents arranged in accordance with a filing plan/system or maintained as a unit because they relate to a particular function or subject, result from the same activity, have a particular form, or because of some other relationship arising out of the circumstances of their creation or use. Also referred to as a record series.

[21] See footnote 20.

## Legislation

The management of this information is defined by specific legislative and policy requirements as well as principles of good management. The legal requirements include:

- the Access to Information Act,
- the Privacy Act,
- the National Archives of Canada Act,
- the National Library Act, and
- the Statistics Canada Act.

These acts describe types of information, protection, access control, and use of the information.

## Policy

Government policy is contained in:

- the Management of Government Information Holdings (MGIH) policy,
- the Security policy,
- the Communications policy,
- the Management of Information Technology policy,
- the Access to Information policy, and
- the Privacy policy.

These policies[23] provide the principles of good management and provide guidance for developping day-to-day procedures.

Information management is an essential function to support, improve, and enhance institutions' ability to fulfil their mission and deliver their programs and services. Managing information as a corporate resource will permit organisations to meet their business needs and to keep and protect the information of continuing value for the future generations. Good information management principles ensure that these goals are met. They also reduce the likelihood of duplication in organisations and control of information holdings and duplication of information collections and hence reduce costs.

---

[22] *Government of Canada Information Management - Responsibilities and Good Practices*. Treasury Board of Canada Secretariat. 1992.

[23] The above policies can be obtained from the Treasury Board Secretariat General Information at 957-2400. A copy of the policies can also be found in most federal governement libraries on the *Treasury Board Secretariat Publications for the Management of the Federal Government on CD-ROM*. April 1996.

As part of good information management practices, information must be understandable, usable and available. In other words, timely access to the correct information comprising the corporate memory of an institution is necessary in the business conduct of an institution. Regardless of the mediums on which the information is recorded, some form of access tool is required. This tool must be able to meet the needs of authorized individuals requiring access to the information. It must be able to point the individual towards the information and, in some cases, provide direct access to the specific information.

The thesaural approach may be an effective strategy for the access and retrieval of a broad range of information, information types, and retrieval criteria. Thesauri can be used to identify and access specialized information in individual data bases or information holdings, as well as provide a general institutional locator system pointing the way to these specific data bases and holdings. Depending on the various business information requirements of the institution, thesauri can be as simple or complex as dictated by the information requirements of the users. A functions thesaurus is a tool which can be used to identify, describe, classify, and organise records and also, to access and retrieve them. Within the context of these applications, a functions thesaurus supports the management of records throughout their life cycle.

Given the exponential increase in the creation of information collections, the use of these collections and the sharing of information between federal government departments, maintaining the quality of information holdings and collections as well as their access over time is a crucial responsibility of institutions.

## Management of Government Information Holdings

*MGIH* policy requires the maintenance of a current, comprehensive and structured identification or classification system or systems which can effectively organize and locate information. This requirement is to help government institutions to plan for, direct, organize and control their information holdings throughout the life cycle of each piece of information. The identification, management, and conservation of information holdings also leads to the ability of government institutions and the public to reconstruct the evolution of policies and programs for historical or archival purposes, or as a part of the identification and protection of essential information.

In order to ensure conformation with information management policies, it is necessary to identify the needs or problems within an institution's information management program to gain an understanding of its strengths and weaknesses and to weigh alternate methods of information management.

## Identifying Information Management Requirements

The steps involved in identifying the organisational requirements include:[24]

1. identifying specific requirements which the institution is not meeting or is meeting in an unsatisfactory manner,
2. providing a broad rationale for any changes to current practices and methods, and
3. preparing and presenting results of the needs identification phase to the appropriate decision makers.

Ultimately, identifying organisational requirements would lead to:

- information availability,
- timely retrieval,
- information protection,
- accountability,
- preservation of information, and
- resource economy.

- The availability of information involves the retrieval of documents, through various mechanisms, for reference as and when required by institution officials for decision-making in support of departmental programs.
- The timely retrieval of information is required to meet requests for information under the Access to Information and Privacy Acts. Loss, deterioration, unwarranted access, and unwarranted destruction are the target activities to be prevented by information protection.
- Accountability ensures that you capture each departmental activity needed to create a corporate information inventory in response to MGIH and other legislative and policy requirements.
- Institutional classification systems must also be linked to the activity structure of the department to be able to report to central agencies.

---

[24] *"Classification Systems Analysis Design and Implementation (CSADI) - Instructors' Manual"*, Module 4: 4.3. (Course offered by National Archives of Canada)

- Preservation of information of archival value is another requirement of good information management in conjunction with the preservation of essential institution information for business resumption as part of disaster recovery planning. Resource economy includes the costs associated with staffing, shelving, floor space, stationary supplies, and technological equipment.

There are a number of reasons why an institution should identify its information management program requirements. These include:

- breakdown in the service provided by the current systems through neglect, a change in departmental activities, a major organization change, a change in legislated requirements, or a change in the types of information being collected and used; and

- signs that the system is breaking down due to problems in information retrieval, classification backlog, user avoidance of the systems, and inconsistent and inaccurate classification and retrieval.

Responding to the following questions may also lead to an identification of further requirements:[25]

- Do standardized procedures for the management of records exist?
- Are multiple systems in use and are they compatible?
- Is there a classification scheme? Indices?
- Are the needs of users being met?
- Do the users use the system?
- Is the information active, semi-active, or inactive?
- Is the information essential?
- Is the information duplicated elsewhere?
- Where is the information physically located?
- Is there sufficient storage space?

---

[25] Ibid. p 4.4.

When all of the requirements for effective information management have been identified, each institution will be in a better position to determine effective strategies for meeting policy requirements and establishing internal procedures, including the identification of the best suited tools to manage its information. They will also be in a better position to operate more effectively and to preserve the corporate memory of the institution.

## 5. FUNCTIONS THESAURUS - WHAT IS IT?

(Prepared by TSWG Sub-Group I: Definition of Requirements)

### Functions Thesaurus

A functions thesaurus[26] is a list of contemporary broad and narrow terms which reflect the major functions and activities carried out by one or several organisations. Functions terms include substantive key terms occurring in legislation, policy and such official documents as financial estimates, budget documents, expenditure plans and other official management and planning documents. The key terms drawn from official documents would be terms describing the mission, the mandate, the programmes, and the activities of the organisation. In other words, a functions thesaurus is constructed using terms which describe the business of the organisation.

What distinguishes a functions thesaurus from any other thesaurus is the source of the indexing terms. In a subject thesaurus indexing terms are drawn from the contents of a document or text; in a functions thesaurus, indexing terms are drawn from official documents of the creating agency. The effect of this difference is **to enable the functions thesaurus to use a business/activity approach which provides contextual information**. On examination, it emerges that business or function terms provide subject or content access as well as business or context access. One may say that a functions thesaurus is a tool for description which incidentally provides subject access.

In addition, a function thesaurus may be defined to include key functional and structural terms as they evolve, and have evolved, over time in official documents; to define, scope, and act as an authority or custodian of function terms in use within an organisation or across organisations[27]; to include organization titles and changes, functions or activities, programmes, and mandates over time; and to map terms both at one point in time and through time. This may be accomplished through the use of scope notes, historical notes, and date ranges. Defined in this way, a functions thesaurus may ensure consistency and act as a reference point for those needing to understand and use terms through their historical evolution. It can also create an audit trail for functions and function words over time.

Two points are to be noted. First, as was stated earlier, a thesaurus is characterised by a hierarchical structure through its broad terms and narrow terms. Second, a thesaurus is characterised by a navigational capacity, that is one can navigate vertically and horizontally through the thesaural structure of broad terms, narrow terms, related terms, and associated terms.

---

[26] Murray-Lachapelle, Rosemary. Creating Intelligent Access to Archival Information on the Internet. A presentation to the Association of Canadian Archivists at the Twentieth Annual Conference. Regina. June 1995.

[27] A functions thesaurus is a tool which can be used across government departments, or across governments.

Due to these two traits, a thesaurus is well suited to index the business functions or activities of organisations which are also characterised by hierarchical and branching structures on vertical and horizontal axes.

The Australian Archives uses a functions thesaurus to manage its records. In this thesaurus, the highest level organisation is the country, Canada, for example, or New Zealand. An example drawn from the Australian Archives would be as follows:

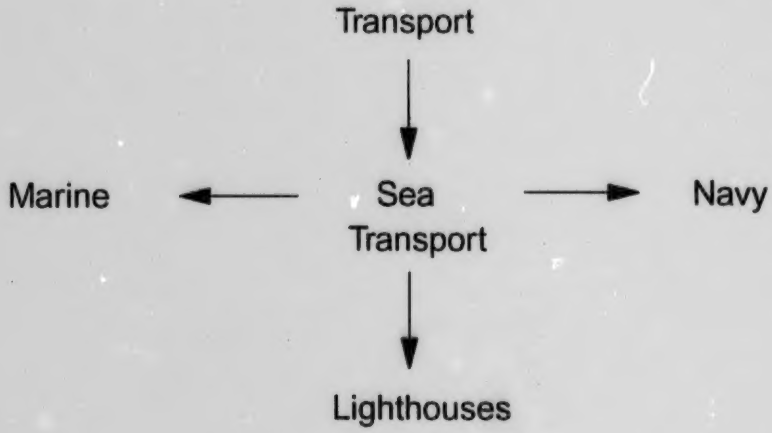
| | |
|---|---|
| ° the heading is: | Transport, |
| ° with a narrow term of: | Sea Transport, and |
| ° a narrower term of: | Lighthouses. |

° Synonyms and related terms may also exist for this term.

As noted on figure 2.1, hierarchy is demonstrated as well as both vertical and horizontal navigational ability within the thesaural structure. If we were looking at a high level organisation chart of the Government of Canada, it would look like figure 2.2.

**Figure 2.1**

# Thesaurus Navigation



Transport

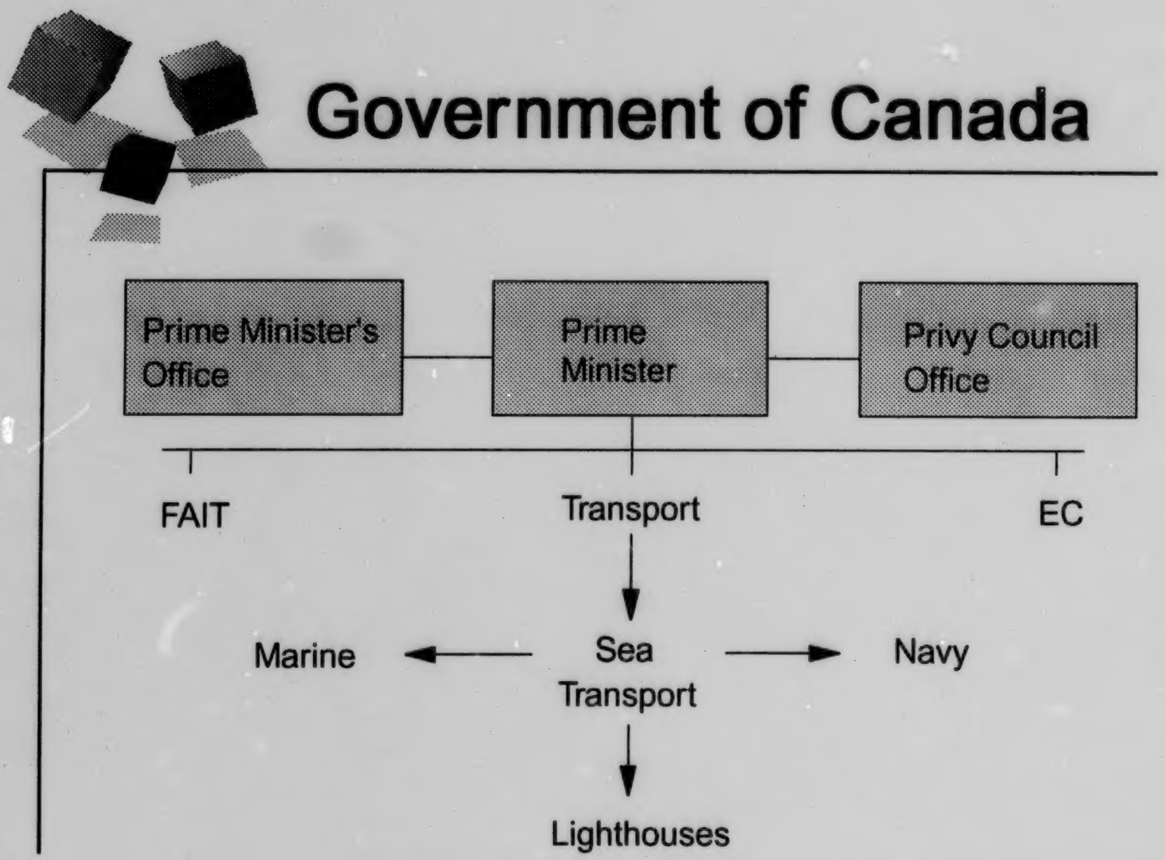
Marine

Sea Transport

Navy

Lighthouses

Figure 2.2


Government of Canada
Prime Minister's Office
Prime Minister
Privy Council Office
FAIT
Transport
EC
Marine
Sea Transport
Navy
Lighthouses

As stated earlier, functions are stable whereas organisations are variable. That is, organisations tend to be periodically reengineered or reorganised, dismembered, renamed and so on. Functions, however, tend simply to be reassigned within or among organisations. Because the indexing terms in a functions thesaurus derive from the "business" or mandate/mission functions (as in the Functions Thesaurus of the Australian Archives,[28]), rather than from a subject typology, functions which are spread among units within an organisation, or among organisations, or related and similar functions held by diverse organisations, may be tracked. As an example: parks may be tracked by national park, by the various federal departments within which it has resided over time, and by the various federal departments within which it may be distributed at the present time. A functions thesaurus, therefore, serves not only as a tool to provide contextual access, but also as a tracking device for functions or business activities.

Search capacities within a functions thesaurus as part of a larger information system, may include search, not only by function, but by organisation (government, department, government official, university, etc., by organisation location (country, city), keywords in the name of the organisation, date range of an organisation, series, etc. Clearly, the elements notes above provide a great deal of contextual information. It is to be noted that the Australian Archives Functions Thesaurus is Commonwealth-wide which demonstrates its value as a mapping device within a diverse organisational structure. The functions approach may be applied in a variety of contexts. The Canadian federal government is one theatre where a functions thesaurus may effectively perform.

A government-wide functions thesaurus can build on what already exists. For example, Infosource uses terms, activities, etc. as they relate to access and privacy programmes for each department. A thesaurus for Infosource, therefore, would provide a common functions thesaurus building block as well as developmental experience in the building process. This is true as well of the government information locator concept which can build upon existing departmental thesauri with established, particular uses in government departments, and upon other building blocks already in existence. The skeletal structure of a government information locator system is already in place. A major next step in the construction of such a government-wide service is to create a government-wide functions thesaurus, that is a thesaurus of functional terms as defined above.

---

[28] Functions Thesaurus. Australian Archives. Rinse Finding Aid No. 2. Information available from the Director, Records Information Management, Australian Archives, P.O. Box 34, Dickson ACT 2602, Australia.

"The FUNCTIONS THESAURUS is a list of contemporary broad and narrow terms which reflect the major functions and activities carried out by commonwealth Government agencies from the Federation to the present day. Functions are allocated to government departments under Administrative Arrangements Orders. Functions move between agencies as a result of changes in the machinery of government.

All Commonwealth Agencies registered in the RINSE database have been allocated one or more terms which describe the main activities, and functions, carried out by them. The dates during which a particular function was carried out by an individual agency is also held on the database." p. 2.

## 6. DEFINITION OF CONCEPTS

(Section prepared by the National Archives of Canada)

This section provides an overview of various thesaural concepts as well as definitions. The definitions are derived from the literature, including the ISO (International Organization for Standardization) standards. The source of the definition is noted in each case.

### Thesaurus

ISO defines thesaurus as "a vocabulary of a controlled indexing language, formally organised so that the a priori relationships between concepts (for example as broader and narrower) are made explicit. Indexing language is a controlled set of terms selected from natural language and used to represent, in summary form the subjects of documents"[29].

A thesaurus is a set of words or phrases, a controlled vocabulary, which has been organised to show the relationships among the words or phrases in order to effectively describe and later retrieve information. The words or phrases are those which best describe the subject area and, when used in relation to one another, best summarize the information. A thesaurus facilitates the management of **concepts** contained in a text. Terms are organised in a hierarchical manner to demonstrate movement from the general to specific. It has also been called the "switching mechanism for reducing the intellectual effort required on the part of the searcher".[30] Thesauri can be displayed in a variety of ways (graphic, tree structure, arrowgraph, systematic displays, or by simple alphabetical listings) and use a variety of symbols and abbreviations the simplest of which would be BT meaning broad term and NT meaning narrow term.

### Thesaurus Characteristics

The construction of a thesaurus begins with an understanding of the collection of information to be searched, regardless of the format of the information, and the information requirements of the organization holding the collection. Some of the questions to be asked include:

- Is the collection in one language only, or is the information available in a number of languages?

- How large is the collection?

---

[29] The International Organization for Standardization. ISO 5964. 1985. p.3.

[30] McLelland, Joe. "Computers, Databases and Thesauri". *Aslib Proceedings*. Vol. 42 no. 7/8. July/August 1990. p. 201-205.

- Does the collection contain only textual documents, or does it contain numeric databases, paper and electronic information, maps, videos, etc? What are the retrieval requirements, for example realtime display, currency, online display, full text display, related information, scope notes, etc?
- How important is the accuracy, the precision and recall factors, of the information retrieved?
- Is more than one database involved requiring a cross referencing of terms?
- What is the best presentation method for this mechanism?

The words or phrases which comprise a thesaurus are called **terms**. When deliberately restricted to precise meanings in order to effectively describe the subject, the terms are known collectively as a controlled vocabulary. Lists of specific authorized terms such as an alphabetical listings of legislative authorities, geographical features, cities of a country, etc. are used when indexing the information contained in the database or subject area of the thesaurus and are known as authority lists or controlled vocabularies.

## Controlled Vocabulary

Controlled vocabulary means specified data values to be used in a system, these data values to be pre-determined. The pre-determined, specified data values are to be used when putting data into data elements in a data structure. Data structures are established by such standards as SGML, ODA/ODIF, and MARC. Data values, or controlled vocabularies including thesauri, are such schemes as subject heading lists, names, geographic regions, functions, or country codes and language codes. Examples would be ISO Country Codes, the Canadian List of Name Authorities, the Canadian Geographic Names Database, and the OECD Macrothesaurus. These lists provide standard spellings, references, symbols, etc. and act as a control mechanism for indexing purposes. Authority lists and controlled vocabularies are available on a wide variety of subjects in hard copy and electronic formats, commercially and through other government departments.

An overall thesaurus-like structure can map to such controlled vocabularies. It is the use of controlled vocabularies, or data values, which are not organised according to a thesaurus structure, within the information system as a whole, which distinguishes the thesaural approach from a thesaurus per se.[31] That is to say that the thesaural approach uses some concepts of thesaurus but without providing the hierarchy (or structure) and relationships among the terms.

---

[31] Strategies and Future Applied R and D Directions for a Corporate-Wide Access Vocabulary. Prepared for Information Holdings Management Branch Environment Canada by Information Management Services INFOMAN Inc. Final Report, March, 1993. Ottawa: Consulting and Audit Canada, 1993.

## Thesaural Approach

By thesaural approach is meant the use of controlled vocabularies government-wide in the construction and management of information systems. The objective is to increase ease of construction by using what already exists, to increase standardisation in order to facilitate the exchange of information among systems and to increase ease of access to the information by information seekers. At a level above the specific information systems, which may use controlled vocabularies and thesauri for search and retrieval, there may be a thesaurus-like retrieval structure. This higher level thesaurus will point to, and initiate negotiations, among systems and databases for purposes of information retrieval.

As an example of the thesaural approach, an authoritative name list (name authority list) is adopted as a standard and made known government-wide. As a result, government departments and agencies when constructing their automated inventories draw from this list when selecting names. Over time this has a standardising effect. The benefits are greater ease and efficiency of information management since the same work (controlled vocabulary) would be shared among departments, and greater effectiveness of retrieval by eliminating dispersion of related information previously entered under different forms of name or subject by different departments or organisational units. The thesaural approach, therefore, promotes the use of controlled vocabularies in building information systems in order to achieve cost savings at the point of system creation and significantly improve information retrieval at the point of system use. The ultimate objective is to effectively identify and describe information in order to make it available, useable, and understandable over time, and to do so in a cost-effective manner.

## Thesaurus Structure

In a standard thesaurus, each term is organized in a hierarchical manner so that the general topics within the subject area are subdivided into increasingly more specific terms. The function of each term is defined and described by a number of abbreviations or symbols which show how the term is to be used, which other terms are related, what synonyms are associated with the term, which terms have broader or narrower meanings, etc. In some cases scope notes are included which further define the term and its use.

It should be emphasized that a thesaurus need not be limited to only textual information holdings or databases. The information formats may be mixed to include video and sound as well as numeric and textual information in fixed fields or free form. Location identifiers may also be appended to individual terms or become the subject of a separate index, electronic or manual.

## Thesaurus Display

The terms of a thesaurus can be organized and displayed in various ways. These include alphabetical listings, systematic displays where terms are grouped into similar subject areas, and graphic displays such as a tree structure or an arrowgraph. A tree structure looks like a typical departmental organization chart with the broadest term appearing at the top of the page and narrower terms printed in descending and expanded order with lines connecting the terms. An arrowgraph starts with a grid where the broadest term is placed in a relatively central position. Narrower terms are arranged around the broadest term, within the grid structure and the levels of detail are indicated by lines and arrows. Related terms are placed outside the grid. Terms can be located by visually consulting the structure or by grid reference.[32] Combinations of these can include a systematic display supported by an alphabetical index of all the terms in the display.

A thesaurus, organized as an alphabetical listing, is generally associated with specific subject areas such as economics, electrical engineering, etc. The number of terms is therefore limited, making it easier for the searcher to look up terms to retrieve the required documents.

A graphic display of terms is also better suited for a small subject area with a small controlled vocabulary. Since the display is a two dimensional picture usually shown on a single page, space limitations generally determine the size of the subject area. Graphic displays can be broken down into a number of more specific displays linked together, but this can create confusion when used as a search tool for retrieval purposes.

A systematic display, on the other hand, when accompanied by an alphabetical listing can be used with large, complex subject areas. A wider range of terms is used with a relatively wider range of relationships between terms.

The most common form of thesaurus display, however, is an alphabetical display with scope notes and indications of inter-term relationships by use of abbreviations or symbols. Some of the abbreviations include SN for scope note, USE to indicate the preferred choice of term when synonyms exist, BT for broad term or the term having the widest meaning, NT for narrow term or the term having the most specific meaning, and RT for related term.[33] The symbols could include location identifiers for the information in the form of database locations, file drawer specification, or system identification.

## Thesaurus Construction/Design

The construction of a thesaurus begins with an understanding of the collection of information to be searched or indexed, regardless of the format of the information. The issues of language, collection size, collection content, retrieval requirements, number of databases involved, and accuracy of retrieval must be addressed first.

---

[32] Description of displays is from *ISO 2788 Guidelines for the Establishment and Development of Monolingual Thesauri*. Examples of displays can be found in *ISO 2788* and *ISO 5964 Guidelines for the Establishment and Development of Multilingual Thesauri*.

[33] See *ISO 2788 Guidelines for the Establishment and Development of Monolingual Thesauri* for approved definitions of abbreviations, and *ISO 5964 Guidelines for the Establishment and Development of Multilingual Thesauri, Annex A* for symbols used in thesaural relationships.

The controlled vocabulary of a thesaurus is the key to its effectiveness. Since the terms are chosen based on their usefulness in describing the information within the subject area and are identified in relation to each other within an entire subject area, the association of concepts and not merely single words is more easily tracked. This means specific information or documents can be more precisely retrieved and the information will be more relevant since whole concepts will have been viewed in relation to each other. The initial choice of thesaural terms and concepts requires consensus and agreement between the various human components of the information storage and retrieval system.

Since thesaural terms and, in a broader context, thesaural concepts are organized in relation to one another, additions and amendments to a thesaurus can be time consuming and require the expertise of a trained indexer who is familiar with the subject area and the existing hierarchy of terms. The automation of this function does, however, decrease the time factor.

### Multilingual Thesaurus

A thesaurus is normally constructed in the spoken language of the searcher and is known as a monolingual thesaurus. However, given the increasing ease with which information is globally collected and exchanged, thesauri are now likely to contain information in several languages. Multilingual thesauri are becoming common place and comprise two or more languages with cross-indexed controlled vocabularies.

Multilingual aspects must also be taken into consideration from the design and throughout the construction stage of a thesaurus. Some of the issues which must be resolved include, for example, how to cross-reference terms between languages, whether singular or plural term variations will be used, non-equivalencies, spelling, and the number of languages involved.

### Retrieval Aspects

The requirements for information retrieval by the users of the information holdings play an important part in determining the construction and nature of a thesaurus. Precision and recall of information by the searcher is a major factor in determining the number of terms used to describe the information and the details associated with each term. **Precision** is the ability of the system not to retrieve irrelevant information. **Recall**, the inverse of this, is the ability of the system to retrieve relevant information. A high degree of precision, the retrieval of specific, relevant information usually results in a low degree of recall, the number of documents retrieved and vice versa.[34]

---

[34] Detailed discussions on the importance of precision and recall can be found in Aitchison, J. and Gilchrist, A. 1987. *Thesaurus Construction: A Practical Manual.* 2d ed. ASLIB. and Lancaster, F.W. 1986. *Vocabulary Control for Information Retrieval.* 2d ed. Virginia:Information Resources Press.

### Term Selection

Term selection can be accomplished in a variety of ways and at various times throughout the creation and life of a thesaurus. Terms can be selected during the creation of the information or from an existing information holding. Terms can be chosen from numerous sources including such standardized terminological forms as published authority lists and controlled vocabularies, existing thesauri, information classification schedules, encyclopaedias, lexicons, dictionaries and glossaries, terminological databases, treatises on the terminology of a subject field, indexes of journals and abstracts and indexes of other publications in the field. As well, terms can be derived by relying on the users' experience and knowledge or the indexers' experience and knowledge or by scanning the existing database or information holding using automated term selection systems. Regardless of the process used to select terms, management needs to assign responsibility to a position or team for the approval of terms selection.

### Thesaurus Maintenance

The maintenance of a thesaurus, regardless of the size or format, manual or automated, requires consistent policy application as well as labour and financial investments. These resources must be dedicated to ongoing thesaurus development and not be viewed as a one time only project. Qualified and experienced personnel are required to meet the needs of hardware maintenance, software support, and the functional requirements of user interface. Although the construction and maintenance of a thesaurus can be very labour and time intensive, the new person interfaces and user friendly software have alleviated some of these problems.

Whether the database or information holding is static or dynamic, uniform indexing of information must be maintained. Terms must be added, deleted and amended on a regular basis to meet the operational requirements of the agency employing the thesaurus and the clients of the agency. The editorial work required to maintain a thesaurus should not be underestimated.

### User Interface

User interface can take the form of direct system query or through qualified intermediaries. If the system is end user operated, the thesaurus should be user friendly, with a minimum of complex devices. The controlled vocabulary should contain uncomplicated terminology and should be transparent to the user. As far as possible, the natural language terms used by the searcher should be cross indexed or mapped to the nearest equivalent controlled terms to improve access through online searching.

### Training

Training, initial and ongoing, is a major time and expense factor in an end user system. Even with user friendly software constantly being developed, the number of searchers requiring training is limited only by the size of the organization and its clients. Training costs, however, are significantly lower when qualified intermediaries, such as a static search only group, is available to the do actual term investigation and identification for the user.

## Automation

Thesauri have been available for a number of years in the areas of medicine and electrical engineering and have traditionally been printed in book or booklet format and updated annually, or less often, through published amendments.

With the advancements in technology today and the growing number of specialized databases, thesauri have become automated in the areas of creation, display and for retrieval purposes. Depending on the requirements of the searcher in the areas of real time display, static or dynamic databases, and size of network, manual thesauri can still be the preferred, although rarely used, format.

A static database is one which contains information that does not change or is not amended in any way; it is merely added to so that the controlled vocabulary remains the same and does not require updating. This is opposed to a dynamic database where the information is constantly changing and being recombined. Real time display of information is most often the preferred form of display in dynamic databases allowing the searcher access to information as it exists at a specific point in time. This is usually the case in databases of journalistic information most often used by news services.

Automation of the thesaural functions provides certain advantages for the end user. One advantage is the flexibility of the format of the thesaural display. Computer programs can represent conceptual information under a number of forms on demand: listings such as alphabetical, hierarchical and thematic, or graphic displays showing the descriptors and their relationships in non-linear formats. Since terms within separate databases can be linked, through automation, greater areas of information can be searched and more relevant information can be retrieved. The flexibility of access is another advantage. In addition to the printed thesaurus, online and interactive programs allow the end user to exploit the linguistic potential of the thesaurus more fully. The end user can move about within the thesaurus and bring into focus more detailed portions of it depending on what is required at a given time.[35]

## Faceted Thesaurus[36]

When creating a thesaurus, the information which will constitute the subject or data base must be analysed and classified in order to determine what terms may be used to build the thesaurus. In systematic classification a sequence of classes of information is brought together and is organized by some type of coding system or notation.[37] One of the simplest types of systematic classification is known as faceted classification. Aitchison and Gilchrist define a faceted classification as "a species of systematic classification that uses facet analysis, a technique involving the allocation of concepts to their fundamental categories, applicable not only in subject field organization, but also in distinguishing different inter-term relationships.... A faceted classification is a type of systematic classification using analytico-synthetic techniques."[38] The analytical part of the technique refers to the way the classification system is structured. The numbers associated with the classes of information are used to represent simple concepts which are organized into clearly defined categories. The synthetic part refers to the ability of the system to combine terms and concepts leading to more complex expressions to describe the information in the data base.

---

[35] Bertrand-Gastaldy, Suzanne and Davidson, Colin H. "Improved Design of Graphic Displays in Thesauri - Through Technology and Ergonomics". *Journal of Documentation.* Vol. 42 no. 4. December 1986.

[36] A detailed discussion and examples can be found in Aitchison, J. and Gilchrist, A. 1987. *Thesaurus Construction: A Practical Manual.* 2d ed. ASLIB.

Facets are essentially subclasses of the original classification system which "produce homogeneous, mutually exclusive groups"[39]. "There is an underlying pattern of fundamental categories in all subject fields, which are represented in the [subclasses] or facets analysed"[40].

When analysing the information data base, general concepts are extracted and divided in broad categories or facets. These concepts are further analysed to determine specific terms which are then grouped into subfacets. This may occur as many times as required in order to retrieve meaningful, useful information.

In thesaurus construction, faceted classification is useful in a number of ways. It "provides a tool for the analysis of subject fields and for determining the relationships between the concepts"[41], as well as being "used as the systematic display in the published thesaurus"[42].

### Common Thesaurus or Metathesaurus

A common thesaurus is one which is shared by more than one organisation. It may be an umbrella-like super structure or meta structure which links to small structures. For examples, areas where massive amounts of information exist, as in large organizations, specialized groups may create small, specific thesauri or controlled vocabulary for their area of interest only. These vocabularies or thesauri are known as microvocabularies or microthesauri. The original application of this process was a specialized subset of terms extracted from and therefore compatible with a larger thesaurus. In other words, a microthesaurus can be defined as a specialized vocabulary that maps to some broader thesaurus and is entirely included within the hierarchical structure of that thesaurus.

---

[37] Aitchison, J. and Gilchrist, A. 1987. *Thesaurus Construction: A Practical Manual.* 2d ed. ASLIB. p 34.

[38] Ibid. p 34.

[39] Ibid. p 50.

[40] Ibid. p 53.

[41] Ibid. p 51.

[42] Ibid.

The broader thesaurus, is, in essence, a generic superstructure to encompass existing vocabularies or microthesauri in diverse subject fields, and is known as a common thesaurus or metathesaurus. The searcher need only review this one metathesaurus in order to have access to many separate information holdings. A variation of this would be the instance where within a government department a common thesaurus is created to map to separate information holdings such as library holdings, record holding, inventory holdings.

The TBITS Thesaurus Standards Working Group recommends the development and use of thesauri, common thesauri, and controlled vocabularies as important building blocks in the construction of a government-wide information locator service which is the ultimate tool needed to produce intelligent access to government information. This might be called a **Government Information Locator Service.**

## Works Cited in the Foregoing Sections

Aitchison, Jean and Gilchrist, Alan. Thesaurus Construction: A Practical Manual. 2d ed. London: The Association for Information Management. 1987.

ASIS '93. Proceedings of the 56th ASIS Annual Meeting, 1993 Volume 30, Columbus, Ohio, October 24-28, 1993. Medford, New Jersey: Published for the American Society for Information Science by Learned Information Inc. 1993.

Bertrand-Gastaldy, Suzanne and Davidson, Colin H. "Improved Design of Graphic Displays in Thesauri - Through Technology and Ergonomics." Journal of Documentation 42, no. 4. 1986.

Blueprint for Renewing Government Services Using Information Technology: Discussion Draft. Ottawa: Treasury Board of Canada, Secretariat. 1994.

Calamai, Peter, "Seeking Knowledge from a Flood of Information". The Ottawa Citizen. August 28, 1993.

Christian, Eliot. Using Z39.50 in an Application for the Government Information Locator Service (GILS). Draft. Washington: United States Geological Survey. 1994.

Classification Systems Analysis Design and Implementation (CSADI) - Instructors' Manual. Module 4: 4.3. Ottawa: National Archives of Canada.

Consolidated Index and Indexing and Access Vocabulary. Thesaurus Functional Requirements Report. Second Draft, March 1994. Prepared by Julia Hudson. (Envirosource Partnership Project). Hull: Environment Canada. 1994.

Corporate Inventory of Information Holdings: Guide to the Structure, Data Model, Data Dictionary. Working Draft. Administrative Policy Branch, Information Management Practices. Ottawa: Treasury Board Secretariat. 1991.

Dictionary of Archival Terminology

Envirosource (TM) Intelligent Interface to Information Holdings, Thesaurus and Consolidated Index, Final Report. Prepared by Consulting and Audit Canada for Environment Canada. (Envirosource Partnership Project). Hull: Environment Canada. 1993.

Functions Thesaurus. Australian Archives. Rinse Finding Aid No. 2. Information available from the Director, Records Information Management, Australian Archives, P.O. Box 34, Dickson ACT 2602, Australia.

Government of Canada Information Management - Responsibilities and Good Practices. Ottawa: Treasury Board Secretariat. 1992.

IMOSA: Information Management and Office Systems Advancement. Overview document. A joint initiative of the National Archives of Canada Government Records Branch and the Department of Communications Canadian Workplace Automation Research Centre (now called Centre for Information Technology Innovation). (IMOSA Report). Ottawa: National Archives of Canada. Revised 1991.

Info Source Specifications for a Bilingual Standards-Compliant Thesaurus. Prepared by Consulting and Audit Canada for the Treasury Board Secretariat. Ottawa: Treasury Board Secretariat. 1994.

International Organization for Standardization. Documentation - Guidelines for the Establishment and Development of Monolingual Thesauri. 2nd ed. Geneva: ISO. 1986. (ISO 2788:1986-11-05)

International Organization for Standardization. Documentation - Guidelines for the Establishment and Development of Multilingual Thesauri. 1st ed. Geneva: ISO. 1985. (ISO 5964:1985-02-15

Lancaster, Frederick William. Vocabulary Control for Information Retrieval. 2d ed. Virginia: Information Resources Press. 1986.

McLelland, Joe. "Computers, Databases and Thesauri." Aslib Proceedings 42, no. 7/8. 1990.

Management of Common Reference Data Definitions. Comptroller General. Ottawa: Treasury Board of Canada. October 27, 1992.

Management of Electronic Documents in the Australian Public Service. A Report prepared by the IESC's Electronic Data Management Sub-committee. Australia: Information Exchange Steering Committee (IESC). 1993.

Managing Information as a Resource. Central Computer and Telecommunications Agency. London: HMSO. 1990.

Prototype Assessment: Thesaurus/Consolidated Index (Enivronment Canada). Prepared by Information Ecosystems Group Inc. for Environment Canada. (Envirosource Partnership Project). Hull: Environment Canada. 1993.

Rapport de l'Étude de faisabilité: Développement d'un modèle pour un thésaurus de base informatisé. Un projet conjoint des Archives nationales du Canada, Direction des documents gouvernementaux et du Ministère des Communications, Centre de recherche sur l'informatisation du travail. (IMOSA Report). Ottawa: National Archives of Canada. 1992.

Strategies and Future Applied R and D Directions for a Corporate-Wide Access Vocabulary. Prepared for Information Holdings Management Branch Environment Canada by Information Management Services INFOMAN Inc. Prepared by Consulting and Audit Canada for Environment Canada. Final Report. (Envirosource Partnership Project). Hull: Environment Canada. 1993.

Thesaurus of Envirosource TM Descriptors Environment Canada Information Holdings. Prepared by Susan Fitzmaurice. (Envirosource Partnership Project). Hull: Environment Canada. 1994. Available from the Environment Canada Library.

# 7. CHECKLIST OF FACTORS TO CONSIDER IN ASSESSING THESAURI

(Section prepared by the National Archives of Canada)

When using information technology to meet corporate and operational information holdings requirements, there are a number of factors which must be taken into consideration. This section provides a list of criteria and questions which may be helpful when evaluating software for use with electronic thesauri. Please note that this list is not exhaustive. The list was developed by Beth Knox of the National Archives of Canada on behalf of the TSWG.

## A. Hardware and Software Requirements

- what standards are being followed? TBITS? Industry? Canadian? American?
- does the hardware/software handle both, French and English character sets? Is the documentation provided in both official languages? Does it have the capacity to add multilingual structures?
- type of software program must be based on needs requirements identified by organization: by type of computer for which developed i.e. microcomputer or mainframe; or by functions to be fulfilled i.e. stand-alone software for the construction and maintenance of thesauri, thesaurus software integrated into retrieval systems.
    - will thesaurus be used for indexing, retrieval or both (this is known as an integrated thesaurus and is recommended for greater consistency control)
- software producer information
- hardware and software requirements
- software installation
- system requirements/system integration
    - can the thesaurus operate on a variety of commonly used technology platforms? (it is important that the thesaurus can operate on the existing, or planned for, operating system within the organization)

    - in creating a thesaurus for a specific electronic database, the thesaurus must be compatible with the operating system of the database i.e. DBMS

- is the thesaurus constructed in a modular way? is it flexible in all aspects both logical and conceptual? can it accommodate information in different formats (multi-media forms)? does it allow for future expansion? is it based on standardised scheme and technical platform?

- vendor support, hotlines, documentation

- the ideal would be to have both an alpha and a graphic display for the same thesaurus. Factors to be considered include cost (i.e. requiring an enhanced graphics package), needs, etc.

- is customization possible? is it required?

## B. Handling and User Interface

- operation and user interface

- is the validation of indexer or user input required i.e. to prevent indexing with unauthorized or misspelled terms?

- can new terms ("candidate terms") be added to the thesaurus during the indexing process?

- is the automatic explosion of searches required?

- training requirements based on user needs (must do a needs analysis of organization to determine usage, types of requests, frequency of use, etc.) plus vendor support in terms of training/documentation (the acquisition of the source code may not be possible or required but complete documentation is necessary)

- is help available i.e documentation form? online help? is there an electronic tutorial? is there a user support hot line?

- how complicated is the menu layering? hotkeys?

- is case sensitivity an issue?

- o is it easy to use? does it provide good mapping, plain language and strong search capabilities?

## C. Functionality and Reliability

- o what standards are being followed i.e. CSA, ANSI, ISO, etc. for thesaurus construction, etc.?

- o functionality: input of terms and classifications; search facilities

    - ▸ how are terms entered for search purposes: by full name, by data element, by code number? is there a prompt, a screen? How many data elements compose a term (i.e. what is the alpha character limit) i.e. 40 characters? Do scope notes automatically appear with the term, are links automatically shown?

    - ▸ Authority lists - available? prepackaged? automatically created? maintenance of? - Maintenance authority approach i.e. can system list all terms and then do relationships or must the updater enter each term and relationship separately? Importance of time and cost factors?

    - ▸ are the vocabularies in indexing and searching compatible?

- o functionality: creation of term relationships

    - ▸ is a list of relationships provided for the user (i.e. narrow term, broad term, etc.)? The minimum provided should be synonymy, hierarchical relation and associative relation

    - ▸ how requested by user i.e. screen, complex syntax, etc?

    - ▸ are node labels or facet indicators available, especially for thesauri for large databases? i.e. term is AUTOMOBILES, facet indicators would break the term down BY SIZE/PRICE, BY MANUFACTURER, BY TYPE, etc.

- o functionality: term relationship validation

    - types of conflict to be aware of and system should correct automatically: i.e. non-postable term (one with a USE relationship to another term) can have no other type of relationship, a term cannot refer to itself, hierarchies cannot be circular, the same pair of terms cannot stand in both the BR/NT and RT/RT relationship to each other

- data structure and limits - relational or purely hierarchically structured? limits on the number of terms that can be accommodated by the system?

- batch functions (involving i.e. different input and output functions): import and export of words/data/structured information? output on different media? different forms of output (ASCII, etc)?

    - output functions can include displaying all standardized lists on the screen, saving as a file or for output on the printer,

- management and system functions (involving i.e. user access control, reorganization of data (single file, all files within database, etc.); maintenance requirements: i.e. knowledgeable staff to perform updating, maintenance, troubleshooting etc., vendor support for both hardware and software

    - control i.e. time, date and personal identifier each time a term is amended/deleted/added?

- user access control: password, restrictions i.e. display/change of screens/functions (menus, error messages, system messages, colours), printing

- statistical package included? For management only or for user needs as well?

- backup routines: automatic? operator initiated?

- management of more than one thesaurus?

- can the thesaurus generate language equivalencies (between English and French, for example) with the capacity to create additional linguistic equivalencies as necessary (as may be the case for law requirements)?

- ability to change to operating system/word processing system without leaving the program?

A number of different printed thesauri are available covering a wide range of topics including geophysical terms, art and architecture, place names, and information science. These are available through the local library networks. Complete electronic thesauri are available for use with standalone systems, or through national and international networks. Some software systems also contain specific thesauri or the program to allow the user to build their own thesauri.

# 8. INVENTORY OF THESAURI, THESAURUS PROJECTS AND CONTROLLED VOCABULARIES IN THE FEDERAL GOVERNMENT

This section contains an inventory of thesauri and controlled vocabularies in the federal government, as well as a listing of thesaurus projects reported as ongoing, being developed or in the planning stage. Information was current at the time the data was collected, in February 1994.

The inventory is the result of a survey of federal government departments and agencies conducted by Christiane Desautels of the National Archives of Canada on behalf of TSWG Sub-group III: Controlled Vocabulary and Thesaural Approach. The intent was to promote sharing of information and experience about thesaurus-related activities, and to reduce duplication of effort.

Activities are organized into one of three categories: Thesauri, Controlled Vocabularies, Thesaurus Projects. The category is that assigned by the respondent on the survey questionnaire. The inventory is followed by a Product/Project Name Index, and a Corporate Name Index to indicate the issuing agency.

Finally, information is provided on the **project team**, the **list of contacts** used to seek responses to the inventory questionnaire and the **questionnaire and cover letter** used for data collection.

## INVENTORY CONTENTS

PROJECT OVERVIEW
  - Project Objectives
  - Methodology
  - Criteria for Inclusion in this Report
  - Constraints
  - Preliminary Analysis

THESAURI
  - AQUATIC SCIENCES AND FISHERIES INFORMATION SYSTEM (ASFIS) THESAURUS
  - ARCHTHE
  - CHILD CARE THESAURUS
  - COAL THESAURUS
  - DESCRIPTIVE CATALOGUING STANDARDS FOR PHILATELIC MATERIAL: THESAURUS
  - DISABILITY RELATED SUBJECT HEADING THESAURUS
  - ENERGY THESAURUS
  - FAMILY VIOLENCE THESAURUS
  - GEOSCAN THESAURUS
  - INFORMATION FILES/PERIODICAL INDEX COLLECTIVES (HEADING) INDEX
  - INFORMATION FILES/PERIODICAL INDEX SUBJECT (HEADINGS) INDEX AND THESAURUS
  - LABOURLINE THESAURUS
  - MACROTHESAURUS FOR INFORMATION PROCESSING IN THE FIELD OF ECONOMIC AND SOCIAL DEVELOPMENT - 4TH ED. (IDRC Version)
  - MINPROC THESAURUS

MINTEC THESAURUS
PARK NOMENCLATURE AND PARKS VISUAL DICTIONARY
PHOCUS THESAURUS
PRECIS
STATISTICS CANADA
THESAURUS ENVIRODOQ
THESAURUS IN LIBRARY AND INFORMATION SCIENCE
THESAURUS OF INFORMATION SCIENCES AND TECHNOLOGIES
THESAURUS OF SUBJECT DESCRIPTORS
WATER RESOURCES THESAURUS

CONTROLLED VOCABULARIES
AGING KEYWORDS
ARCHIVAL TERMINOLOGY
BANK OF CANADA LIST OF SUBJECT HEADINGS
BIOGRAPH
CANADIANA AUTHORITIES
CANADIAN GEOGRAPHICAL NAMES DATABASE (CGNDB)
CANADIAN INVENTORY OF HISTORIC BUILDING (CIHB)
CANADIAN SUBJECT HEADINGS (CSH)
CANDITH
CAN/MARC COUNTRY CODES
CAN/MARC GEOGRAPHIC AREA CODES
CAN/MARC LANGUAGE CODES
CATHES
CIPC
CONFERENCES AUTHORITY LIST
DEPOSITORY SERVICES PROGRAM LIST OF SUBJECT DESCRIPTORS
DIS

INFORMATION FILES/PERIODICAL INDEX INDIVIDUALS (HEADING) INDEX
INFORMATION SERVICES GLOSSARY
INTEGRATED STAFFING SYSTEM ABBREVIATION LIST/STANDARD TERMS FOR NAMING ELEMENTS AND SYSTEM COMPONENTS
LIBRARY OF CONGRESS SUBJECT HEADINGS (LCSH)
MANAGEMENT KEYWORD LIST
NAMES AUTHORITY LIST
NAVAL CONSTRUCTION, 16TH - 18TH CENTURIES
ORGANIZATIONS AUTHORITY LIST
POSTS AUTHORITY LIST
SUBJECT HEADING LIST FOR BIBLIOGRAPHIES DATABASE
SUBJECT HEADING LIST FOR FOREIGN AFFAIRS STATEMENTS AND SPEECHES/NEWS RELEASES DATABASE
TERMIUM SUBJECT-FIELD GUIDE/RÉPERTOIRE DES DOMAINES TERMIUM
THE MANUSCRIPT DIVISION SUBJECT HEADING LIST (working title)
THE MANUSCRIPT DIVISION NAME AUTHORITY (working title)
THESMAS
TOPONYMY & TERMINOLOGY IN USE IN THE CANADIAN SERVICE OF PARKS
TREATIES AUTHORITY LIST
TRILINGUAL SPECIES NAME LIST FOR THE WAVES/VAGUES DATABASE

THESAURUS PROJECTS

- CANADIAN HERITAGE CORPORATE THESAURUS
- ENVIROSOURCE
- INFOSOURCE
- MINISIS THESAURUS MANAGEMENT SYSTEM
- STANDARD DEPARTMENTAL SUBJECTS PROJECT
- TERMINI
- VERTICAL FILE THESAURUS

PRODUCT NAME INDEX

INDEX OF CORPORATE NAMES

SUB-GROUP III MEMBER LIST

CONTACT LIST

COVER LETTER AND QUESTIONNAIRE

# PROJECT OVERVIEW

## Project Objectives

The inventory is intended to benefit those organizations eager to develop tools to enhance their ability to describe and preserve records in context, and to retrieve them in a cost-effective manner.

Other objectives of the inventory are:

- to share information;
- to share experience; and
- to identify common sources and structures of information in order to avoid duplication of effort across the federal government.

Ultimately, it is hoped that the inventory will assist departments and agencies in managing their information, in building inventories of information holdings as well as common systems.

## Methodology

A questionnaire was developed to gather data on thesauri, thesaurus projects and controlled vocabularies in the Canadian federal government. It was intended to gather information of general interest for government-wide use. A cover letter enclosed with the questionnaire described the purpose and benefits of building the inventory. They were both circulated to members of Sub-group III for their input and were then tested by a number of participants. (See the questionnaire and cover letter in Appendix C.)

An initial list of contacts was created over the summer 1993 while conducting preliminary work for the survey. The contact list included government officials involved with some aspect of thesauri, thesaurus projects, or controlled vocabularies within their organisation. The list was eventually enhanced by members of the TSWG who provided additional contact names. (See the contact list in Appendix B.)

Individuals on the contact list were reached by telephone to seek their participation in building the inventory by completing a brief questionnaire. A cover letter and questionnaire were subsequently sent by fax and completed questionnaires were returned to the project team by fax. This process enabled the project team to reach several participants and to obtain the required information in a short period of time.

The information gathered from the questionnaires is presented in this report in the form of a brief narrative description for each information product (i.e. thesaurus, thesaurus project or controlled vocabulary). The section entitled "Report Organisation" (which precedes the table of contents) explains how the information has been organised.

By distributing this working document, the project team hopes to receive comments and suggestions from the TSWG members. Please address your comments/suggestions to Rosemary Murray-Lachapelle at the National Archives of Canada at (613) 947-1513, or by fax at (613) 947-1500.

### Criteria for Inclusion in this Report

To be included in this report, an information product had to fit the description of a thesaurus or of a controlled vocabulary as outlined in the cover letter (see Appendix C). Although the study was focussed on information products in use in the Canadian federal government in the National Capital Region, we did not exclude those from outside the region when they were identified to us.

The cover letter also indicates that "this questionnaire is not intended to inventory commercial retrieval services using thesauri unless such a service is used to index the institutions' own information holdings".

### Constraints

In the questionnaire, the "available from" field in question one means where can the actual thesaurus or controlled vocabulary be acquired from. However, it might have been interpreted by certain respondents as referring to where can the "TH/CV software/hardware" (which is the previous field) be acquired from.

The meaning of "E/F Bilingual" field in question two of the questionnaire had not been clearly defined and may have been subject to various interpretations. For example, does bilingual mean that the thesaurus is interactive in French and English, or does it mean that two separate versions exist, one in French and one in English? In the narrative description of the inventory, the language is reported as per indicated in questionnaires.

Although efforts were made to reach the largest possible number of individuals responsible for thesauri, thesaurus projects and controlled vocabularies in the National Capital Region, it is possible that some may have been omitted. We would appreciate notification if any omissions are noticed.

The usefulness and value of this inventory will be assessed in the light of deciding whether to update the inventory periodically.

## Summary of Results

The following statistics were derived from the survey results.

- Of the 38 controlled vocabularies listed, five are available on microform, twenty six are available in electronic format, twenty nine are available in paper format, and nineteen are available in both formats. Results also indicate that only two of those available in electronic format are on CD ROM.
- Of the twenty four thesauri reported, two included less than five hundred terms, one included less than one hundred terms, and sixteen included over one thousand terms. In two cases the number of terms was unknown by the participant and in one case the question was left blank.
- Of the thirty six controlled vocabularies inventoried, six include less than five hundred terms, seven included less than one thousand terms, and nineteen included over one thousand terms. In five cases the question was left blank.
- A total of ten thesauri out of twenty four were reported to be bilingual, while a total of nineteen out of thirty eight controlled vocabularies were reported as such. It should be noted that these statistics may be somewhat askew due to the ambiguity of the term "E/F bilingual".
- Of the twenty four thesauri reported, eighteen were created in-house, four were created by an outside body and two were not reported upon. Comparatively, the controlled vocabulary results indicate that thirty three out of thirty eight were created in-house, four were created by an outside body and one was not reported upon.
- Of the thesauri, fifteen were created prior to 1988 and eight were created since 1988. Comparatively, twenty one controlled vocabularies were created prior to 1988 and sixteen created since 1988.
- In most cases thesauri and controlled vocabularies are used for both indexing and retrieving information. Twenty four out of thirty eight controlled vocabularies are used in an online database and seventeen out of twenty four thesauri are also used online.
- Only three respondents reported that the thesaurus is used to help manage and control an inventory of information holdings.
- Eight out of twenty four thesauri are based on the ISO standards.

## AQUATIC SCIENCES AND FISHERIES INFORMATION SYSTEM (ASFIS) THESAURUS

**Major subject(s):** Aquatic natural & applied sciences; marine, fishery, ship & fish food technologies; living and non-living resources exploitation & processing; aquatic pollution; aquatic environmental changes conservation, public health; and social, economic and policy relevant aspects
**Language(s):** English

**Developed by:** United Nations Food and Agriculture Organisation
**Date created:** 1986
**Derived from:** Thesaurus of Terms Aquatic Sciences Fisheries/FAO, 1976

**Maintained by:** United Nations Food and Agriculture Organisation
**Latest revision:** Underway
**Updating frequency:** As required

**Format(s):** Paper
**Software/hardware used:** SPINES
**Size:** 6,500 descriptors and 3,400 non-descriptors
**Standard(s):** Follows a facet grouping initiated pre-1976

**Uses:** For indexing and retrieving data in bibliographic records and abstracts

**Available from:** United Nations
**Contact:** Fisheries and Oceans
Heather Cameron
**Phone:** (613) 993-2926
**Fax:** (613) 990-4901

## ARCHTHE

**Major subject(s):** Archivistique/gestion des documents
**Minor subject(s):** Conservation
**Language(s):** Bilingue: anglais/français

**Developed by:** Centre canadien d'information et de documentation en archivistique (CCIDA), Archives nationales du Canada
**Date created:** 1987

**Maintained by:** Archives nationales
**Updating frequency:** Au besoin

**Software/hardware used:** Revelation
**Size:** 3,252 termes

**Used by:** Personnel du CCIDA
**Number of users:** 3
**Uses:** Pour l'indexation et le repérage de données sur une base de données en-direct

**Available from:** Archives nationals du Canada
**Contact:** Yves Marcoux
**Phone:** (613) 996-7686
**Fax:** (613) 943-8491
**Internet:** ymarcoux@archives.ca

## CHILD CARE THESAURUS

**Major subject(s):** Child day care
**Minor subject(s):** Child development
**Language(s):** Bilingual: English/French

**Developed by:** Departmental Library, Health Canada
**Date created:** ca 1989

**Maintained by:** Health Canada
**Updating frequency:** As required

**Format(s):** Paper, ASCII, or WordPerfect

**Used by:** Library staff and clients
**Uses:** For indexing and retrieving bibliographic records in an online database using the Sydney Micro Library System software

**Comments:** This thesaurus, along with the relevant bibliographic collection, will be transferred to Human Resources and Development as a result of government reorganisation. The transfer is expected for January 1995.

**Available from:** Health Canada
**Contact:** M. Lovelock
**Phone:** (613) 957-1547
**Fax:** (613) 941-8007

## COAL THESAURUS

**Type:** Keyword index
**Major subject(s):** Coal science
**Minor subject(s):** Coal industry, environment, coal usage, and coal mining
**Language(s):** English

**Developed by:** IEA Coal Research
**Date created:** 1978

**Maintained by:** IEA Coal Research
**Latest revision:** 1994
**Updating frequency:** As required

**Format(s):** Paper and CD-ROM

**Number of users:** 5
**Uses:** To index and retrieve data in abstracts and free/full text in an online database

**Available from:** The vendor: CANMET-Library and Documentation Services Division (Canadian rep. of IEA Coal Research)
**Contact:** Natural Resources Canada
Marc Nason
**Phone:** (613) 992-8837
**Fax:** (613) 952-2587

## DESCRIPTIVE CATALOGUING STANDARDS FOR PHILATELIC MATERIAL: THESAURUS

**Major subject(s):** Philately
**Language(s):** English

**Developed by:** Canadian Postal Archives, National Archives of Canada (NA)
**Date created:** 1990

**Maintained by:** National Archives of Canada
**Latest revision:** 1990
**Updating frequency:** As required

**Format(s):** Paper and WordPerfect
**Size:** 200 terms

**Used by:** Canadian Postal Archives
**Uses:** In an online database with the MINISIS software

**Comments:** A small, highly specialized thesaurus developed to provide standard terminology for the description of philatelic (postage stamp) related material.

**Available from:** National Archives of Canada
**Contact:** Cimon Morin
**Phone:** (613) 996-1199
**Fax:** (613) 992-3744
**Internet:** cmorin@archives.ca

## DISABILITY RELATED SUBJECT HEADINGS THESAURUS

**Type:** Subject heading list
**Major subject(s):** Disability
**Minor subject(s):** Social policy issues
**Language(s):** Bilingual: English/French

**Developed by:** Canadian Clearinghouse on Disability Issues (CCDI)
**Date created:** In process
**Derived from:** "Thésaurus: personne handicapée", LCSH and others

**Maintained by:** Canadian Clearinghouse on Disability Issues
**Updating frequency:** Semi-annually or as required

**Size:** ca 1,500 terms (includes not used terms)

**Used by:** CCDI
**Uses:** To index and retrieve bibliographic records; to provide a possible source of subject authorities for contributors of bibliographic records to a national database on disability publications

**Comments:** At the moment we are using a catalogue on CD ROM provided by Bibliofile, but we will also be using this subject heading list as part of an online catalogue on INNOPAC.

**Available from:** Canadian Clearinghouse on Disability Issues
**Contact:** Shayna Keces
**Phone:** (613) 994-7514
**Fax:** (613) 953-4797

## ENERGY THESAURUS

**Type:** Keyword index
**Major subject(s):** Energy technology (includes no economic terms)
**Minor subject(s):** All forms of fossil and renewable nuclear energy
**Language(s):** English

**Developed by:** Energy Technology Data Exchange (IEA)
**Date created:** ca 1988

**Maintained by:** Energy Technology Data Exchange (IEA)
**Latest revision:** 1994
**Updating frequency:** As required

**Format(s):** Paper and ASCII
**Size:** 5,000 terms

**Number of users:** 5
**Uses:** To index and retrieve data in abstracts and free/full text in an online database

**Available from:** CANMET Library and Documentation Services Division
**Contact:** Natural Resources Canada
Marc Nason
**Phone:** (613) 992-8837
**Fax:** (613) 952-2587

## FAMILY VIOLENCE THESAURUS

**Type:** Subject heading list
**Major subject(s):** Family violence
**Language(s):** Bilingual: English/French

**Developed by:** Health Canada National Clearinghouse on Family Violence
**Date created:** 1988

**Maintained by:** Departmental Library, Health Canada.
**Updating frequency:** As required

**Format(s):** Paper and ASCII

**Used by:** Library staff and its clients

**Uses:** For indexing and retrieving bibliographic records in an online database using the Sydney Micro Library System software

**Available from:** Health Canada
**Contact:** M. Lovelock
**Phone:** (613) 957-1547
**Fax:** (613) 941-8007

## GEOSCAN THESAURUS

**Type:** Authority file and code list
**Major subject(s):** Geoscience
**Language(s):** English

**Developed by:** Geoscan participating agencies

**Maintained by:** Geological Survey of Canada
**Latest revision:** September 1990
**Updating frequency:** As required

**Format(s):** Paper, WordPerfect and ASCII
**Size:** 4,296 terms

**Used by:** Database builders and searchers
**Number of users:** 14
**Uses:** To index and retrieve data in bibliographic records

**Comments:** We use CV's to also allow us to expand codes into French and/or English on output, but also for searching in some cases. The thesaurus of keywords allows for thesaurus-aided searching (NT, BT, RT).

**Available from:** Geological Survey of Canada/Natural Resources Canada
**Contact:** Bruce Blair
**Phone:** (613) 996-9502
**Fax:** (613) 943-8742

## INFORMATION FILES/PERIODICAL INDEX COLLECTIVES (HEADING) INDEX

**Type:** Authority file and subject heading list
**Major subject(s):** Music (organisations, institutions, ensembles)
**Language(s):** 45% English, 30% French and 25% bilingual

**Developed by:** Music Division, National Library of Canada (NLC)
**Date created:** 1970s (card file)

**Latest revision:** May 1996

**Format(s):** Paper and LAN computer file. WWW version available on the NL Web site (http://www.nlc-bnc.ca) as of September 1996
**Software/hardware used:** Microsoft Word (for paper and LAN versions)
**Size:** 2,500 terms
**Standard(s):** Canadiana Authorities and Encyclopedia of Music in Canada

**Used by:** Users and staff of the NLC Music Division; and Internet users of the Canadian Music Periodical Index
**Number of users:** 50 +/ month (for paper and LAN versions)
**Uses:** To index and retrieve free/full text, data in bibliographic records in an online database using the Inmagic Plus software; to help manage and control a departmental inventory of information holdings

**Comments:** This index includes the corporate name of each organisation, with relationships and relevant information: place, date frame, subject classification.

**Available from:** National Library of Canada
**Contact:** Florence Hayes
**Phone:** 947-2702
**Fax:** 952-2895

## INFORMATION FILES/PERIODICAL INDEX SUBJECT (HEADINGS) INDEX AND THESAURUS

**Type:** Authority file, a subject heading list and a lexicon/dictionary
**Major subject(s):** Music (broad subjects, e.g. jazz, instrument-building)
**Language(s):** Bilingual (English and French)

**Developd by:** Music Division, National Library
**Date created:** 1970s card file

**Maintained by:** Music Division, National Library
**Latest revision:** September 1993
**Updating frequency:** As required

**Format(s):** Paper and LAN computer file. WWW version available on the NL Web site (http://www.nlc-bnc.ca) as of September 1996
**Software/hardware used:** Microsoft Word (for paper and LAN versions)
**Size:** 700 terms
**Standard(s):** Canadiana SH, LCSH, Music Index, and Encyclopedia of Music in Canada

**Used by:** Users and staff of the NLC Music Division; and Internet users of the Canadian Music Periodical Index
**Number of users:** 50+/ month (for paper and LAN versions)
**Uses:** To index and retrieve free/full text and data in bibliographic records in an online database using Inmagic Plus software; to help manage and control a departmental inventory of information holdings

**Comments:** This index-thesaurus includes the established subject heading, its relationships and a scope note to define the term or give other relevant information.

**Available from:** National Library of Canada
**Contact:** Florence Hayes
**Phone:** 947-2702
**Fax:** 952-2895

## LABOURLINE THESAURUS

**Major subject(s):** Industrial relations, labour
**Language(s):** English and French

**Developed by:** Labour Canada Library Services in conjunction with the Committee on Industrial Relations Librarians
**Date created:** September 1981

**Maintained by:** Labour Canada Library Services
**Latest revision:** 1991
**Updating frequency:** Annually

**Format(s):** Paper
**Software/hardware used:** MINISIS

| | |
|---|---|
| **Size:** | Approximately 8,000 terms |
| **Standard(s):** | Presently being converted to meet ISO standards |
| **Used by:** | Labour Canada Library |
| **Uses:** | To index and retrieve bibliographic records in an online database |
| **Available from:** | Human Resources Development Canada |
| **Contact:** | Helen Apouchtine |
| **Phone:** | 953-0028 |
| **Fax:** | 953-2098 |

## MACROTHESAURUS FOR INFORMATION PROCESSING IN THE FIELD OF ECONOMIC AND SOCIAL DEVELOPMENT (IDRC Version)

**Major subject(s):** Economic and social development
**Language(s):** Trilingual: English, French, and Spanish

**Developed by:** Organisation for Economic Co-operation and Development (OECD) in Paris. The International Development Research Centre (IDRC) is a member of the international Macrothesaurus Advisory Committee and has played a major role in the development of this thesaurus.

**Date created:** The Macrothesaurus was created in 1972. The IDRC version has been in use since 1974.

**Maintained by:** OECD maintains the core vocabulary in English, French and Spanish. The IDRC Library is responsible for ensuring the IDRC version of the Macrothesaurus is updated to meet its needs.
**Latest revision:** 4th edition 1991; IDRC version updated in August 1992
**Updating frequency:** Normally, once a year for the IDRC version. In early 1995, 30 new terms will be added.

**Format(s):** Available in printed and machine-readable formats (ISO-2709 and ASCII) from OECD, Paris)
**Software/hardware used:** Micro-CDS/ISIS thesaurus management software (OECD); MINISIS software (IDRC). MINISIS is used for both thesaurus management and for online searching of the associated bibliographic and research-activity databases.
**Size:** 5,000 terms
**Standard(s):** ISO 2788, ISO 5964

**Used by:** IDRC Library database (BIBLIOL); IDRC's database of research activities (RADIUS); Inter-Agency Development Research Information System (IDRIS), which is managed by IDRC. The IDRC version is also used by the International Development Information Centre at the Canadian International Development Agency (CIDA).

**Available from:** International Development Research Centre
**Contact:** Maureen Sly
**Phone:** (613) 236-6163, extension 2575
**Fax:** (613) 238-7230
**Internet:** msly@idrc.ca

## MINPROC THESAURUS

**Type:** Keyword index
**Major subject(s):** Physical metallurgy
**Minor subject(s):** Hydrometallurgy, pyrometallurgy, ore benefaction, chemical metallurgy, equipment, health and safety and environment
**Language(s):** Bilingual: English, French

**Developed by:** CANMET Library and Documentation Services Division (LDSD)
**Date created:** ca 1980

**Maintained by:** Library and Documentation Services Division
**Latest revision:** 1994
**Updating frequency:** As required

**Format(s):** Paper
**Size:** 1,000

**Used by:** CANMET-LDSD
**Number of users:** 5
**Uses:** To index and retrieve data in abstracts and free/full text in an online database

**Available from:** CANMET - LSDS, priced publication
**Contact:** Natural Resources Canada
Mac Nason
**Phone:** (613) 992-8837
**Fax:** (613) 952-2587

## MINTEC THESAURUS

**Type:** Keyword index
**Major subject(s):** Mining engineering
**Minor subject(s):** Health and safety, environment, mine stability, rock mechanics, mining equipment and rock breakage
**Language(s):** Bilingual: English, French

**Developed by:** CANMET Library and Documentation Services Division (LDSD)
**Date created:** ca 1980

**Maintained by:** Library and Documentation Services Division
**Latest revision:** 1994

**Updating frequency:** As required

**Format(s):** Paper
**Size:** 1,000

**Used by:** CANMET - LDSD
**Number of users:** 5
**Uses:** To index and retrieve data in abstracts in an online database

**Available from:** Natural Resources Canada
**Contact:** Mac Nason
**Phone:** (613) 992-8837
**Fax:** (613) 952-2587

## PARK NOMENCLATURE AND PARKS VISUAL DICTIONARY

**Type:** Authority file and lexicon/dictionary
**Major subject(s):** Artifacts at National Historic Sites (historic objects & reproductions)
**Language(s):** Bilingual: English/French

**Developed by:** Parks Canada, National Historic Sites Interpretation Branch
**Date created:** 1984
**Derived from:** Chenwall Nomenclature

**Maintained by:** Parks Canada
**Latest revision:** Ongoing
**Updating frequency:** As required

**Format(s):** Paper
**Software/hardware used:** Incorporated in Artifact Information System (AIS) support tables
**Standard(s):** ISO

**Used by:** Parks and many museums
**Uses:** To index and retrieve data in an online database which is being redeveloped. A structured naming system for all objects in Parks Historic Objects Collection.

**Available from:** Canadian Heritage/Parks Canada

**Contact:** Rosemary Campbell
**Phone:** 993-0632
**Fax:** 990-6627

## PHOCUS THESAURUS

**Major subject(s):** Photographic materials conservation
**Minor subject(s):** Paper conservation
**Language(s):** English

**Developed by:** Picture Conservation Division of the National Archives of Canada (NA)
**Date created:** ca 1982

**Maintained by:** National Archives
**Updating frequency:** As required

**Format(s):** Paper
**Software/hardware used:** MINISIS
**Size:** 1,000 terms

**Used by:** Staff and researchers
**Number of users:** 30
**Uses:** To index and retrieve data in bibliographic records in an online database

**Available from:** National Archives of Canada
**Contact:** Alison Hale
**Phone:** (613) 996-9280
**Fax:** (613) 943-8491
**Internet:** ahale@archives.ca

## PRECIS

PRECIS is the name of the thesaural system itself and is the name used by the National Archives for the thesaurus they have developed.

**Major subject(s):** Broad subject range - covers any subject which may be shown on maps, etc. Also includes some architectural terms. Lot of geographical names.
**Language(s):** English and French

**Developed by:** National Archives of Canada, Visual and Sound Archives

**Maintained by:** National Archives of Canada, Visual and Sound Archives
**Latest revision:** Continuous
**Updating frequency:** Continuous

**Format(s):** Paper, microform or tape
**Software/hardware used:** ISM/UTLAS data base
**Size:** 8,501 terms (This count does not include the strings, just the thesaural terms)
**Standard(s):** ISO

**Used by:** Visual and Sound Archives Division staff
**Uses:** To retrieve bibliographic records in an online database; we also have a microfiche catalogue - one part of which is for subject

**Comments:** PRECIS is a thesaurally based subject indexing system which consists of two parts: the thesaurus and the subject strings.

**Available from:** National Archives of Canada
**Contact:** Velma Parker
**Phone:** (613) 996-7611
**Fax:** (613) 995-6575

## STATISTICS CANADA THESAURUS

**Major subject(s):** Drawn from Statistics Canada publications terminology - a vast array of subjects
**Language(s):** English and French

**Developed by:** Statistics Canada Library (SCL)
**Date created:** 1988

**Maintained by:** Statistics Canada
**Latest revision:** March 1989
**Updating frequency:** As required

**Format(s):** Paper and WordPerfect
**Standard(s):** ISO

**Used by:** SCL
**Uses:** To index and retrieve bibliographic records and abstracts

**Comments:** It is hoped that it will soon be applied to an automated environment.

**Available from:** Statistics Canada
**Contact:** Fay Hjartarson
**Phone:** (613) 951-0953
**Fax:** (613) 951-0939

## THÉSAURUS DES SCIENCES ET DES TECHNOLOGIES DE L'INFORMATION

**Major subject(s):** Technologies de l'information et des communications
**Minor subject(s):** Informatique; télécommunications; optique; acoustique; traitement de l'information; traitement des langues naturelles
**Language(s):** Bilingue: français/anglais

**Developed by:** Service intégré des ressources d'information au Centre d'innovation en technologie de l'information, Industrie Canada
**Date created:** 1991

**Maintained by:** Centre d'innovation en technologie de l'information
**Latest revision:** 1994: en ligne

**Format(s):** Papier et Internet
**Software/hardware used:** MINISIS/HP 3000
**Size:** 12,411 termes
**Standard(s):** ISO

**Uses:** Pour l'indexation et le repérage de données dans des champs sur une base de données en-direct; également utilisé comme outil de référence

**Available from:** Groupe Communication Canada
**Contact:** Centre d'innovation en technologie de l'information
Industrie Canada
Anne Simard
**Phone:** (514) 973-5746
**Fax:** (514) 973-5757

## THESAURUS ENVIRODOQ

**Major subject(s):** Water resources
**Minor subject(s):** Environment

**Language(s):** French

**Date created:** May 1984
**Format(s):** Paper
**Size:** Approximately 200

**Uses:** To index bibliographic records in an online database using the Inmagic Plus software

**Contact:** Environment Canada
Jean Bennett
**Phone:** (819) 953-9427
**Fax:** (819) 994-0237

## THESAURUS IN LIBRARY AND INFORMATION SCIENCE

**Major subject(s):** Library and information science
**Minor subject(s):** Literacy and publishing
**Language(s):** English with partial translation into French

**Developed by:** National Library of Canada (NLC)
**Date created:** 1971 (manual file), thesaurus was automated in 1993

**Maintained by:** National Library of Canada (until 1995)
**Latest revision:** CEASED as of 1995
**Updating frequency:** CEASED as of 1995

**Format(s):** Paper, Microsoft Word and ASCII
**Software/hardware used:** Term Manager
**Size:** 1,900 descriptors and 800 cross-references
**Standard(s):** ISO, partially

**Used by:** Staff and clients
**Uses:** To index and retrieve data in bibliographic records in an online database using the INMAGIC Plus software

## THESAURUS OF SUBJECT DESCRIPTORS

**Major subject(s):** Foreign relations, diplomacy, and international trade
**Minor subject(s):** Environment, defence

**Language(s):** Bilingual: English/French

**Developed by:** Foreign Affairs and International Trade (FAIT)/MKRC
**Date created:** 1978

**Maintained by:** Foreign Affairs and International Trade
**Latest revision:** Ongoing
**Updating frequency:** As required

**Format(s):** ASCII (KWIC available in paper)
**Software/hardware used:** Basis/Basisplus
**Size:** 6,500 terms
**Standard(s):** ISO

**Used by:** Records management staff and CATS (Corporate Automated Text Storage System) users
**Number of users:** 3,000
**Uses:** To index and retrieve data in fields in an online database

**Available from:** Foreign Affairs and International Trade
**Contact:** Diane Crouse
**Phone:** (613) 996-9385
**Fax:** (613) 943-1090

## WATER RESOURCES THESAURUS

**Major subject(s):** Economic, legal, social, engineering, recreational, biological, geographical, ecological, and qualitative aspects of water resources
**Language(s):** English

**Developed by:** U.S. Department of the Interior, Office of Water Research & Technology
**Date created:** 1971

**Derived from:** Selected water resources abstracts
**Maintained by:** U.S. Department of the Interior, Office of Water Research & Technology
**Latest revision:** ca 1980

**Format(s):** Paper
**Size:** 2,000 terms
**Standard(s):** ISO

**Uses:** To index bibliographic records in an online database using the Inmagic Plus software

**Available from:** Department of Interior, Office of Water Research & Technology, Technology & Information Transfer Program, Washington, D.C. 20240)
**Contact:** Environment Canada
Jean Bennett
**Phone:** (819) 953-9427
**Fax:** (819) 994-0237

## CONTROLLED VOCABULARIES

### AGING KEYWORDS

**Type:** Subject heading list
**Major subject(s):** Gerontology and geriatrics
**Language(s):** Bilingual: English/French (currently being "bilingualized")

**Developed by:** Departmental Library of Health Canada
**Date created:** 1989
**Derived from:** *Thesaurus of Aging Terminology* (U.S.)

**Maintained by:** Health Canada
**Updating frequency:** As required

**Format(s):** Paper and ASCII

**Used by:** The library staff and clients
**Uses:** To index and retrieve bibliographic records in an online database, the software of which is Sydney Microlibrary System

**Available from:** Health Canada
**Contact:** Marty Lovelock
**Phone:** (613) 957-1547
**Fax:** (613) 941-8007

### ARCHIVAL TERMINOLOGY

**Type:** Lexicon/dictionary
**Major subject(s):** Archival terminology
**Language(s):** Bilingual: English/French

**Developed by:** Office of Archival Standards (OAS), National Archives of Canada (NA)
**Date created:** November 1992

**Maintained by:** OAS and a departmental committee
**Latest revision:** November 1992
**Updating frequency:** As required

**Format(s):** Paper format
**Software/hardware used:** WHOTERM

**Used by:** NA staff
**Uses:** For all archival work

**Comments:** All terms include definitions

**Available from:** National Archives of Canada
**Contact:** Hugo Stibbe
**Phone:** (613) 996-7592
**Fax:** (613) 995-2267
**Internet:** hstibbe@archives.ca

## BANK OF CANADA LIST OF SUBJECT HEADINGS

**Type:** Subject heading list
**Major subject(s):** Banking, economics and finance
**Minor subject(s):** Management and electronic data processing
**Language(s):** English. Has a French index

**Developed by:** Library of the Bank of Canada (BC)
**Date created:** 1935
**Derived from:** LC, CSH and SLA (Banking and Financial Subject Headings - Special Libraries Association)

**Maintained by:** Bank of Canada Library
**Latest revision:** March 1, 1994
**Updating frequency:** As required

**Format(s):** Paper and WordPerfect
**Size:** 2,000 - 2,500 terms

**Used by:** Bank of Canada staff
**Uses:** For retrieving bibliographic records; to help manage and control a departmental inventory of information holdings; used on MICROISIS to maintain index of Bank of Canada publications

**Available from:** Bank of Canada
**Contact:** Suzanne Le Blanc
**Phone:** (613) 782-8786
**Fax:** (613) 782-8655

## BIOGRAPH

**Type:** Authority file
**Major subject(s):** Canadian artists and photographers
**Minor subject(s):** Other Canadian individuals, corporations and organisations
**Language(s):** Bilingual: English/French

**Developed by:** National Archives of Canada
**Date created:** 1989

**Maintained by:** National Archives
**Latest revision:** 1994
**Updating frequency:** As requested

**Format(s):** ASCII format
**Software/hardware used:** MINISIS/HP 3000
**Size:** 18,500 terms

**Used by:** National Archives Visual & Sound Archives Division
**Number of users:** 20
**Uses:** In an online database

**Available from:** National Archives of Canada
**Contact:** Gerald Stone
**Phone:** (613) 996-7790
**Fax:** (613) 995-6226
**Internet:** gstone@archives.ca

## CANADIAN GEOGRAPHICAL NAMES DATABASE (CGNDB)

**Type:** Authority file
**Major subject(s):** Geographical names in Canada
**Minor subject(s):** Attribute fields - latitude, longitude, NTS map no., province, etc.
**Language(s):** English or French

**Developed by:** The Canadian Permanent Committee on Geographical Names (CPCGN), Natural Resources Canada (NRC)
**Date created:** 1897

**Maintained by:** Natural Resources Canada (NRC)
**Updating frequency:** Ongoing

**Format(s):** Paper, microform and ASCII
**Software/hardware used:** Oracle/Unix/Sun
**Size:** 500,000 names

**Used by:** Names used on Canadian maps; supplied to enquirers - publishers, researchers, librarians, etc.
**Uses:** An authority source; to index and retrieve field data in an online database, which uses Oracle software

**Comments:** The CGNDB is a Canada-wide database of geographical names and attribute data. It was developed from card records of the CPCGN and its predecessors - starting with names made official in 1897. Data is listed in the "Gazettes of Canada Series" and can be purchased in digital form.

**Available from:** Natural Resources Canada
**Contact:** Helen Kerfoot
**Phone:** (613) 992-3405
**Fax:** (613) 943-8282

## CANADIAN INVENTORY OF HISTORIC BUILDINGS (CIHB)

**Type:** Code list, a keyword index, a subject heading list and various indexes
**Major subject(s):** Built heritage/inventory of buildings
**Minor subject(s):** Policies, planning, management, designation lists, etc.
**Language(s):** Bilingual: English/French (some in French or English only)

**Developed by:** Parks Canada
**Date created:** 1970

**Maintained by:** Parks Canada
**Latest revision:** 1989-90
**Updating frequency:** As required

**Format(s):** Paper, microform, ASCII and WordPerfect, and partial CD ROM
**Software/hardware used:** Foxpro/Foxbase

**Used by:** Researchers; Historic Sites & Monuments Board of Canada; Heritage Groups; and the public
**Number of users:** 600/annual (average)

| | |
|---|---|
| **Uses:** | To index and retrieve field and abstract data and to retrieve bibliographic records in an online database |
| **Comments:** | The CIHB system is based on the geocode (modified) for key access |
| **Available from:** | Canadian Heritage |
| **Contact:** | Richard Martineau |
| **Phone:** | (819) 994-2867 |
| **Fax:** | (819) 953-4909 |

## CANADIAN SUBJECT HEADINGS (CSH)

| | |
|---|---|
| **Type:** | Subject heading list |
| **Major subject(s):** | All topics of Canadian content or interest |
| **Language(s):** | English. It is linked to French-language subject headings in the Répertoire de vedettes-matières, produced by Université Laval. |
| **Developed by:** | National Library of Canada (NLC) |
| **Date created:** | 1978 |
| **Maintained by:** | National Library |
| **Latest revision:** | 3rd edition, published in 1992 |
| **Updating frequency:** | Updated with a cumulating semi-annual supplement |
| **Format(s):** | Paper |
| **Size:** | 6,000 terms and 25,000 entry vocabulary |
| **Standard(s):** | It is modeled on the Library of Congress Subject Headings |
| **Used by:** | NLC and other Canadian libraries |
| **Uses:** | To index and retrieve bibliographic records; to provide access to topics in Canadiana |
| **Comments:** | CSH is a precoordinated controlled vocabulary list specially designed to be used in tandem with LCSH to offer in depth subject retrieval fro Canadian library materials, both print and non-print. |
| **Available from:** | Canada Communications Group |
| **Contact:** | National Library of Canada<br>Alina Schweitzer |
| **Phone:** | 953-6810 |
| **Fax:** | 953-0291 |

## CANADIANA AUTHORITIES

**Type:** Authority file
**Major subject(s):** Corporate and personal names used on publications of Canadian origin or interest, including standardized headings for Canadian federal and provincial government bodies, treaties, laws and conferences
**Language(s):** Bilingual: English/French

**Developed by:** National Library of Canada
**Date created:** 1975

**Maintained by:** National Library of Canada
**Latest revision:** Ongoing
**Updating frequency:** Cumulating quarterly base file with cumulating bi-weekly supplements

**Format(s):** Microform and magnetic tape or tape cartridge and also available online
**Size:** Over 430,000 standardized headings
**Standard(s):** Canadian MARC Communication Format: Authorities; and Anglo-American Cataloguing Rules
**Used by:** Mainly libraries and documentation centres
**Number of users:** 450 microfiche subscribers, 11 magnetic tape subscribers and 650 online users

**Uses:** To index and retrieve free/full text; data in bibliographic records and abstracts in an online database; to help manage and control a departmental inventory of information holdings.

**Comments:** Canadiana authorities include English and French forms of name headings complete with their associated cross-references. History notes provide additional information on corporate bodies.

**Available from:** Bibliographic Services, National Library of Canada
**Fax:** (819) 997-7517
**Internet:** canadiana@nlc-bnc.ca

## CANDITH

**Type:** Descriptor list
**Major subject(s):** Defence-related terms in the following areas: aeronautics, applied sciences, communications, computer science, engineering, medical science, military science and physical science
**Language(s):** English

**Developed by:** Lotus CSG Canada Ltd. for National Defence (ND)
**Date created:** March 1993
**Derived from:** Thesaurus of Engineering and Scientific Terms and the Canadian Supplement

**Maintained by:** Directorate Scientific Information Services of ND
**Latest revision:** Ongoing
**Updating frequency:** As required

**Format(s):** Paper
**Software/hardware used:** Basisplus/TM

**Used by:** Staff of Directorate
**Number of users:** 16
**Uses:** To index and retrieve bibliographic records in an online database; to help manage and control a departmental inventory of information holdings

**Available from:** National Defence
**Contact:** Rota Bouse
**Phone:** 992-2294
**Fax:** 996-0392

## CAN/MARC COUNTRY CODES

**Type:** Code list
**Major subject(s):** Country codes
**Language(s):** English and French

**Developed by:** National Library of Canada (NLC)
**Date created:** 1973
**Derived from:** USMARC code list for countries

**Maintained by:** National Library
**Latest revision:** January 1996
**Updating frequency:** As required

**Format(s):** Paper
**Size:** 360 codes

**Used by:** Canadian libraries and archives
**Uses:** For retrieval of bibliographic records and data in fields

**Comments:** Published as an appendix to the Canadian MARC Communication Format for Bibliographic Data and to the Format de communication du MARC canadien pour les données bibliographiques

**Available from:** Canada Communication Group - Printing
**Contact:** National Library of Canada
Young-Hee Queinnec
**Phone:** 994-6936
**Fax:** 953-8508

## CAN/MARC GEOGRAPHIC AREA CODES

**Type:** Code list
**Major subject(s):** Geographic area codes
**Language(s):** English and French

**Developed by:** National Library of Canada (NLC)
**Date created:** 1973
**Derived from:** USMARC code list for geographic areas

**Maintained by:** National Library of Canada
**Latest revision:** January 1994
**Updating frequency:** As required

**Format(s):** Paper
**Size:** 533 codes

**Used by:** Canadian libraries and archives
**Uses:** For the retrieval of bibliographic records and data in fields

**Comments:** Published as an appendix to the Canadian MARC Communication Format for Bibliographic Data and to the Format de communication du MARC canadian pour les données bibliographiques.

**Available from:** Canada Communication Group - Printing

**Contact:** National Library of Canada
Young-Hee Queinnec
**Phone:** 994-6936
**Fax:** 953-8508

## CAN/MARC LANGUAGE CODES

| | |
|---|---|
| **Type:** | Code list |
| **Major subject(s):** | Language codes |
| **Language(s):** | English and French |
| **Developed by:** | National Library of Canada (NLC) |
| **Date created:** | 1973 |
| **Derived from:** | USMARC code list for languages |
| **Maintained by:** | National Library of Canada |
| **Latest revision:** | January 1994 |
| **Updating frequency:** | As required |
| **Format(s):** | Paper |
| **Size:** | 372 codes |
| **Standard(s):** | ANSI Z39.53 |
| **Used by:** | Canadian libraries and archives |
| **Uses:** | For indexing and retrieving data in bibliographic records in an online database using Inmagic Plus |
| **Comments:** | Published as an appendix to the Canadian MARC Communication Format for Bibliographic Data and to the Format de communication du MARC Canadien pour les données bibliographiques |
| **Available from:** | Canada Communication Group - Printing |
| **Contact:** | National Library of Canada<br>Young-Hee Queinnec |
| **Phone:** | 994-6936 |
| **Fax:** | 953-8508 |

## CATHES

| | |
|---|---|
| **Type:** | Authority file |
| **Major subject(s):** | Defence-related science and technology corporate names |
| **Language(s):** | English |
| **Developed by:** | Lotus CSG Canada Ltd. for the National Defence (ND) |
| **Date created:** | March 1993 |
| **Derived from:** | DSIS Corporate Source Authority List |

**Maintained by:** Directorate Scientific Information Services (DSIS) of ND
**Latest revision:** Ongoing basis
**Updating frequency:** As required

**Format(s):** Paper and online
**Software/hardware used:** Basisplus/TM
**Size:** 10,000 terms

**Used by:** Directorate staff
**Number of users:** 16
**Uses:** To index and retrieve data in bibliographic records in an online database using Basisplus; to help manage and control a departmental inventory of information holdings

**Available from:** National Defence
**Contact:** Rota Bouse
**Phone:** 992-2294
**Fax:** 996-0392

## CONFERENCES AUTHORITY LIST

**Type:** Authority file
**Major subject(s):** Names of conferences relating to foreign affairs
**Language(s):** Bilingual: English/French

**Developed by:** Foreign Affairs and International Trade (FAIT)
**Date created:** 1978

**Maintained by:** Foreign Affairs and International Trade
**Latest revision:** Ongoing
**Updating frequency:** As required

**Format(s):** ASCII
**Software/hardware used:** BASIS/Basisplus
**Size:** 2,000 terms
**Standard(s):** ISO

**Used by:** Records management staff and CATS (Corporate Automated Text Storage System) users
**Number of users:** 3,000
**Uses:** For indexing and retrieving data in an online database

**Available from:** Foreign Affairs and International Trade
**Contact:** Diane Crouse
**Phone:** (613) 996-9385
**Fax:** (613) 943-1090

## CPIC (CANADIAN POLICE INFORMATION CENTRE)

**Type:** Code list and keyword index
**Language(s):** Bilingual

**Developed by:** CPIC office of the Royal Canadian Mounted Police
**Date created:** 1972

**Maintained by:** CPIC office of the Royal Canadian Mounted Police
**Latest revision:** Yearly
**Updating frequency:** As required

**Used by:** Canadian Police Community
**Number of users:** 2,500 accredited police agencies
**Uses:** To index data in an online database using the CPIC software

**Available from:** Royal Canadian Mounted Police
**Contact:** Sgt. Rod Berg
**Phone:** (613) 998-7599
**Fax:** (613) 952-7889

## DEPOSITORY SERVICES PROGRAM LIST OF SUBJECT DESCRIPTORS

**Type:** Subject heading list
**Major subject(s):** General fields of knowledge for Canadian Government Documents
**Language(s):** Bilingual: English/French

**Developed by:** Depository Services Program (DSP), Canada Communication Group Publishing
**Date created:** January 1993

**Maintained by:** Depository Services Program (DSP), Canada Communication Group Publishing
**Updating frequency:** As required

F2

**Format(s):** Paper
**Software/hardware used:** MINISIS
**Size:** 1,000 terms
**Standard(s):** ISO

**Used by:** Presently in-house, targeted for marketing and sales personnel and DSP clients

**Uses:** For indexing and retrieving data in bibliographic records in an online database; to help manage and control a departmental inventory of information holdings.

**Available from:** Canada Communication Group - Publishing
**Contact:** Public Works and Government Services Canada
Judy Patterson, Depository Services Program
**Phone:** (819) 997-5366
**Fax:** (819) 956-6341

## DIS

**Type:** Subject heading
**Major subject(s):** Library services to persons with disabilities
**Minor subject(s):** Disabilities; alternate format publications; adaptive technology
**Language(s):** English

**Developed by:** National Library of Canada (NLC)
**Date created:** 1983
**Derived from:** ABLEDATA Thesaurus, Australian Disability Thesaurus, CRCD (Canadian Rehabilitation Council for the Disabled) Rehabilitation Classification Scheme, DLF (Disabled Living Foundation) Thesaurus, REHABDATA Thesaurus, and the subject headings used in the NLS Reference Collections (Library of Congress) National Library Service for the Blind and Physically Handicapped).

**Maintained by:** National Library of Canada (until 1995)
**Latest revision:** CEASED as of 1995
**Updating frequency:** CEASED as of 1995

**Format(s):** Paper and Microsoft Word
**Software/hardware used:** Microsoft Word
**Size:** 456

**Used by:** NLC staff and patrons
**Uses:** For indexing and retrieving data in bibliographic records in an online database using Inmagic Plus

**Contact:** Reference and Information Services, National Library of Canada
**Phone:** 995-9481
**Fax:** 943-1112
**Internet:** reference@nlc-bnc.ca

## INFORMATION FILES/PERIODICAL INDEX INDIVIDUALS (HEADING) INDEX

**Type:** Authority file and subject heading list
**Major subject(s):** Music (performers and all individuals involved in the field of music)
**Language(s):** English 65%, French 35%

**Developed by:** National Library (NLC), Music Division
**Date created:** In process February 1994
**Maintained by:** National Library
**Updating frequency:** As required

**Format(s):** Paper
**Software/hardware used:** Microsoft Word software for eventual transfer to Inmagic Plus
**Size:** ca 3,200
**Standard(s):** Canadiana Authorities and the Encyclopedia of Music in Canada

**Used by:** Staff/users of NLC Music Division
**Number of users:** ca 50/per month
**Uses:** To index and retrieve free/full text (information files), data in fields (subject fields) and bibliographic records (periodical index) in an online database using the Inmagic Plus software in 1994-95

**Comments:** This index includes the established name of an individual with relationships (see, see also references) and potentially will include birth/death dates and other relevant information, e.g. subject classification.

**Available from:** National Library of Canada
**Contact:** Florence Hayes
**Phone:** 947-2702
**Fax:** 952-2895

## INFORMATION SERVICES GLOSSARY

**Type:** Glossary
**Major subject(s):** Definitions on communications, information and data
**Language(s):** English; being translated (1994)

**Developed by:** Directorate Information Systems Architecture of the National Defence
**Date created:** March 1994
**Derived from:** Various NATO documents

**Maintained by:** National Defence
**Latest revision:** 1st edition

**Format(s):** Paper and WordPerfect
**Size:** 670 definitions
**Standard(s):** NATO and DND - A.DP040 publication

**Used by:** ND information management staff
**Number of users:** Approximately 1,000-1,500
**Uses:** To help manage and control a departmental inventory of information terms

**Comments:** DND is trying to establish common terminology.

**Available from:** National Defence
**Contact:** Bob Dowell
**Phone:** 996-3903
**Fax:** 996-6277

## INTEGRATED STAFFING SYSTEM ABBREVIATION LIST/ STANDARD TERMS FOR NAMING ELEMENTS AND SYSTEM COMPONENTS

**Type:** Authority file
**Major subject(s):** Staffing/person
**Minor subject(s):** Language assessment/testing
**Language(s):** English

**Developed by:** Informatics Directorate/ISS Project, Public Service Commission (PSC)
**Date created:** 1992

**Maintained by:** Public Service Commission
**Latest revision:** April 28, 1994
**Updating frequency:** As required

**Format(s):** Paper, WordPerfect 5.1
**Software/hardware used:** WordPerfect
**Size:** 700

**Used by:** PSC ISS project
**Number of users:** 75

**Uses:** For indexing free/full text in an online database using the INGRES software; to help manage and control a departmental inventory of information holdings

**Available from:** Public Service Commission
**Contact:** Richard Harrington
**Phone:** (613) 992-3324
**Fax:** (613) 943-2030

## LIBRARY OF CONGRESS SUBJECT HEADINGS (LCSH)

**Type:** Subject heading list
**Major subject(s):** General and universal list of headings for subject retrieval of library materials
**Language(s):** English

**Developed by:** Library of Congress (LC)
**Date created:** ca 1900

**Maintained by:** Library of Congress
**Latest revision:** 17th ed., 1994
**Updating frequency:** Annually

**Format(s):** Paper and microform
**Size:** 200,000 entry vocabulary
**Standard(s):** It follows the Library of Congress Subject Cataloguing Manual

**Used by:** NLC and most other Canadian institutions for subject access to their collections, in conjunction with Canadian Subject Headings which provides more detailed and in-depth subject treatment for Canadian topics

**Uses:** To index and retrieve bibliographic records

**Comments:** Largest and most universally-used list of precoordinated controlled retrieval terms for library materials in English (also adapted into French (Répertoire de vedettes-matières) and other languages.

**Available from:** Library of Congress

**Contact:** National Library of Canada
Alina Schweitzer
**Phone:** 953-6810
**Fax:** 953-0291

## MANAGEMENT KEYWORD LIST (now called MANAGEMENT RESOURCE CENTRE THESAURUS)

**Type:** Keyword list
**Major subject(s):** Management development and executive development
**Language(s):** English and French

**Developed by:** Management Resource Centre (MRC) of the Canadian Centre for Management Development (CCMD)
**Date created:** April 1993

**Maintained by:** Canadian Centre for Management Development
**Latest revision:** May 1994
**Updating frequency:** As required

**Format(s):** Paper and ASCII
**Software/hardware used:** Inmagic
**Size:** 1300

**Used by:** CCMD SearchMagic System
**Number of users:** 125
**Uses:** For indexing and retrieving bibliographic records in an online database using the SearchMagic software

**Available from:** Canadian Centre for Management Development
**Contact:** Hesch Hanley
**Phone:** (613) 992-9355
**Fax:** (613) 995-0331

## MANUSCRIPT DIVISION NAME AUTHORITY (working title)

**Type:** Subject heading list and name authority list
**Major subject(s):** Names relating to the content of our archival holdings at the general fonds/collection level
**Language(s):** English - will be translated when completed

**Developed by:** M & A Resources for the National Archives of Canada
**Date created:** 1990
**Derived from:** Canadiana and a local authority devised by J. Cumming

**Maintained by:** National Archives of Canada
**Latest revision:** 1994
**Updating frequency:** Monthly

**Format(s):** Paper and WordPerfect
**Software/hardware used:** MINISIS
**Size:** 200 pages
**Standard(s):** ISO

**Used by:** Archivists of the Manuscript Division of National Archives
**Number of users:** 25
**Uses:** To index and retrieve free/full text documents and bibliographic/descriptive records in an online database

**Comments:** This is a name authority devised while indexing all our fonds/collection descriptions (inventories). This is also a specific field in our database C930. This information has been migrated to ARCHIVIA.

**Available from:** National Archives of Canada - Manuscript Division
**Contact:** Diane Beattie
**Phone:** (613) 996-7291
**Fax:** (613) 943-8112
**Internet:** dbeattie@archives.ca

## MANUSCRIPT DIVISION SUBJECT HEADING LIST (working title)

**Type:** Subject heading list
**Major subject(s):** Relate to the content of our archival holdings at the general fonds/collection level

**Language(s):** English - will be translated when completed

**Developed by:** M & A Resources for the National Archives of Canada (NA)
**Date created:** 1990
**Derived from:** Library of Congress Subject Headings/Canadian Subject and a local authority by J. Cumming

**Maintained by:** National Archives
**Latest revision:** 1994
**Updating frequency:** Monthly

**Format(s):** Paper and WordPerfect
**Software/hardware used:** MINISIS
**Size:** 300 pages
**Standard(s):** ISO

**Used by:** Archivists of the Manuscript Division, NA
**Number of users:** 25
**Uses:** To index and retrieve free/full text documents and bibliographic/descriptive records in an online database

**Comments:** This subject authority was devised while indexing all of our fonds/collection descriptions (i.e. inventories). This information appears in a specific subject field C940 in the MINISIS database. These descriptions have also been migrated to a CD-ROM application called ARCHIVIA which provides full text search capabilities.

**Available from:** National Archives of Canada
**Contact:** Diane Beattie
**Phone:** (613) 996-7291
**Fax:** (613) 943-8112
**Internet:** dbeattie@archives.ca

## NAMES AUTHORITY LIST

**Type:** Authority file
**Major subject(s):** Names of persons active in some aspect of foreign relations
**Language(s):** English/French bilingual

**Developed by:** Foreign Affairs and International Trade (FAIT)
**Date created:** 1978

**Maintained by:** Foreign Affairs and International Trade
**Latest revision:** Ongoing
**Updating frequency:** As required

**Format(s):** ASCII
**Software/hardware used:** Basisplus software
**Size:** 83,000 terms
**Standard(s):** ISO

**Used by:** Records management staff and the users of the Corporate Automated Text Storage System
**Number of users:** 3,000
**Uses:** For indexing and retrieving data in an online database

**Available from:** Foreign Affairs and International Trade
**Contact:** Diane Crouse
**Phone:** (613) 996-9385
**Fax:** (613) 943-1090

## NAVAL CONSTRUCTION, 16TH - 18TH CENTURIES

**Type:** Authority file
**Major subject(s):** Naval construction, 16th-18th centuries
**Language(s):** Multilingual: English, French, Portuguese, Spanish

**Developed by:** Parks Canada - History Directorate and Archaeology Directorate
**Date created:** 1986

**Maintained by:** Parks Canada
**Latest revision:** 1994
**Updating frequency:** As required

**Format(s):** Paper
**Size:** Approximately 3000 terms

**Used by:** Parks Canada - Archaeology
**Number of users:** 30-100
**Uses:** For indexing free/full text; to help manage and control a departmental inventory of information holdings.

**Comments:** Vocabularies were developed as historical research tools for textual analysis of 16th - 18th century materials; in turn these vocabularies served to identify and interpret archaeological findings.

This information is not organised for use by other agencies, nor will it be in 1994-95.

**Available from:** Canadian Heritage/Parks Canada
**Contact:** Robert Grenier
**Phone:** 993-2470
**Fax:** 952-1756

## ORGANISATIONS AUTHORITY LIST

**Type:** Authority file
**Major subject(s):** Names of organisations involved in aspects of foreign relations
**Language(s):** Bilingual: English/French

**Developed by:** Foreign Affairs and International Trade (FAIT)
**Date created:** 1978

**Maintained by:** Foreign Affairs and International Trade
**Latest revision:** Ongoing
**Updating frequency:** As required

**Format(s):** ASCII
**Software/hardware used:** BASIS/Basisplus
**Size:** 15,500
**Standard(s):** ISO

**Used by:** Records management staff and users of the CATS (Corporate Automated Text Storage) System
**Number of users:** 3,000
**Uses:** For indexing and retrieving data in an online database

**Available from:** Foreign Affairs and International Trade
**Contact:** Diane Crouse
**Phone:** (613) 996-9385
**Fax:** (613) 943-1090

## POSTS AUTHORITY LIST

**Type:** Authority file and code list
**Major subject(s):** Names of Canadians in other countries' missions
**Language(s):** Bilingual: English/French

**Developed by:** Foreign Affairs and International Trade (FAIT)
**Date created:** 1978

**Maintained by:** Foreign Affairs and International Trade
**Latest revision:** Ongoing
**Updating frequency:** As required

**Format(s):** ASCII
**Software/hardware used:** BASIS/Basisplus
**Size:** 2,500
**Standard(s):** ISO

**Used by:** Records management staff and users of the CATS (Corporate Automated Text Storage) system
**Number of users:** 3,000
**Uses:** For indexing and retrieving data in an online database

**Available from:** Foreign Affairs and International Trade
**Contact:** Diane Crouse
**Phone:** (613) 996-9385
**Fax:** (613) 943-1090

## RÉPERTOIRE DES DOMAINES TERMIUM

**Type:** Liste de codes et liste de vedettes-matières
**Major subject(s):** Tous
**Language(s):** Bilingue: anglais/français

**Developed by:** Université de Montréal pour Travaux publics et services gouvernementaux Canada (TPSGC)
**Date created:** 1976
**Derived from:** Library of Congress

**Maintained by:** Travaux publics et services gouvernementaux Canada
**Latest revision:** 1988
**Updating frequency:** Au besoin

**Format(s):** Papier et WordPerfect
**Software/hardware used:** Basisplus
**Size:** 3,300 termes

**Used by:** Direction de la terminologie de TPSGC
**Number of users:** 100
**Uses:** Pour l'indexation de fiches terminologiques et pour le repérage de documents et de données sur une base de données en-direct

**Comments:** Le VC sert à organiser et repérer des fiches dans la base de données terminologiques multilingues, TERMIUM. Le contenu de cette base de données pourrait lui-même servir comme point de départ pour la création de vocabulaires et de thésaurus bilingues.

**Available from:** Travaux publics et services gouvernementaux Canada
**Contact:** Christine Leonhardt
**Phone:** (819) 997-7929
**Fax:** (819) 994-3670

## SUBJECT HEADING LIST FOR BIBLIOGRAPHIES DATABASE

**Type:** Authority file and a subject heading list
**Major subject(s):** Canada foreign/trade/economic policy, NAFTA, FTA, world policy, peace and security
**Minor subject(s):** Environment, international organisations and international business
**Language(s):** English

**Developed by:** Foreign Affairs and International Trade (FAIT)
**Date created:** 1989
**Maintained by:** Foreign Affairs and International Trade
**Latest revision:** 1993
**Updating frequency:** As required

**Format(s):** Paper, WordPerfect
**Software/hardware used:** WordPerfect
**Size:** 300

**Used by:** The reference librarian and bibliographer
**Uses:** For indexing and retrieving bibliographic records in a database using Nota Bene/Ibid; to help manage and control a departmental inventory of information holdings

**Comments:** Used for flagging documents for retrieval for inclusion in subject bibliographies on Canada foreign/trade/economic/ public policy, etc. This vocabulary list permits rapid indexing for production of bibliographies required by library reference clients, etc. Includes cross-references (see, see also).

**Available from:** Foreign Affairs and International Trade
**Contact:** David Milne
**Phone:** (613) 996-5399
**Fax:** (613) 944-0222

## SUBJECT HEADING LIST FOR FOREIGN AFFAIRS STATEMENTS AND SPEECHES/NEW RELEASES DATABASE

**Type:** Authority file and a subject heading list
**Major subject(s):** Canada foreign policy, trade policy (NAFTA, FTA), international relations and foreign aid
**Minor subject(s):** Environment, security, human rights and international organisations
**Language(s):** Bilingual: English/French

**Developed by:** Foreign Affairs and International Trade (FAIT)
**Date created:** 1988

**Maintained by:** Foreign Affairs and International Trade
**Latest revision:** 1993
**Updating frequency:** As required

**Format(s):** Paper and WordPerfect
**Software/hardware used:** WordPerfect 5.2
**Size:** 500 terms in each language

**Used by:** The reference librarian and bibliographer
**Uses:** For indexing and retrieving bibliographic records in an online database using the software Inmagic; to help manage and control a departmental inventory of information holdings

**Comments:** We were unable to find an existing thesaurus or list that met our needs, namely, specialization in the field of foreign affairs with a subject range extending over all international issue areas. Includes geographic names (countries, regions). Also includes cross-references (see, see also) and some scope notes.

**Available from:** Foreign Affairs and International Trade
**Contact:** David Milne
**Phone:** (613) 996-5399
**Fax:** (613) 944-0222

### THESMAS

**Type:** Authority file
**Major subject(s):** General, social sciences and humanities.
**Language(s):** Bilingual

**Developed by:** National Archives of Canada (NA)
**Date created:** 1985
**Derived from:** LCSH

**Maintained by:** National Archives
**Latest revision:** 1993
**Updating frequency:** As required

**Format(s):** Paper
**Software/hardware used:** Minisis/HP 3000

**Size:** 40,000 terms
**Standard(s):** ISO, ANSI

**Used by:** Visual and Sound Archives Division, National Archives of Canada
**Number of users:** 10
**Uses:** To index and retrieve bibliographic records in an online database

**Available from:** National Archives of Canada
**Contact:** Gerald Stone
**Phone:** (613) 996-7790
**Fax:** (613) 995-6226
**Internet:** gstone@archives.ca

### TOPONYMIE ET TERMINOLOGIE EN USAGE À PARCS CANADA

**Type:** Index des mots-clés
**Major subject(s):** Toponymes des parcs, lieux et canaux
**Minor subject(s):** Message de sécurité, signalisation
**Language(s):** Bilingue: français/anglais

**Developed by:** Parcs Canada
**Size:** 500 termes
**Standard(s):** ISO; le Comité permanent canadien des noms géographiques, CUENCO; et le Répertoire toponymique du Québec

**Used by:** Employés, traducteurs, secteurs privés
**Number of users:** 1,000 et plus
**Uses:** Pour l'indexation et le repérage des données dans des champs

**Comments:** Il y a un projet de bulletin de terminologie spécifique à Parcs Canada en préparation en collaboration avec le Service gouvernemental de terminologie. Nos deux documents seront disponibles via le «TERMIUM» de la Direction de terminologie des services gouvernementaux.

**Contact:** Patrimoine canadien - Parcs Canada
Clément Bédard
**Phone:** 994-5126
**Fax:** 994-5140

## TREATIES AUTHORITY LIST

**Type:** Authority file
**Major subject(s):** Names of treaties
**Language(s):** Bilingual: English/French

**Developed by:** Foreign Affairs and International Trade (FAIT)
**Date created:** 1978

**Maintained by:** Foreign Affairs and International Trade
**Latest revision:** Ongoing
**Updating frequency:** As required

**Format(s):** ASCII
**Software/hardware used:** Basisplus
**Size:** 850 terms
**Standard(s):** ISO

**Used by:** Records management staff and users of the CATS (Corporate Automated Text Storage) system
**Number of users:** 3,000
**Uses:** For indexing and retrieving data in an online database

**Available from:** Foreign Affairs and International Trade
**Contact:** Diane Crouse
**Phone:** (613) 996-9385
**Fax:** (613) 943-1090

## TRILINGUAL SPECIES NAME LIST FOR THE WAVES/VAGUES DATABASE

**Type:** Subject heading list
**Major subject(s):** Taxonomic names of fish, shellfish, etc.
**Minor subject(s):** English and French common names for above

**Developed by:** Library Services Branch, Fisheries and Oceans (DFO), Ottawa
**Date created:** 1986

**Maintained by:** Library, Pacific Biological Station, DFO, Nanaimo, B.C.
**Latest revision:** 1991
**Updating frequency:** As required

**Format(s):** Paper and WordPerfect
**Software/hardware used:** PC File, Version 5
**Size:** 2,100 entries with associated cross-references
**Standard(s):** Follows the DFO WAVES/VAGUES database standards

**Used by:** DFO libraries
**Number of users:** 20
**Uses:** To index and retrieve bibliographic records

**Comments:** Additional information may be obtained through the Library

**Available from:** ButtonWave
**Contact:** Fisheries and Oceans
Heather Cameron
**Phone:** (613) 993-2926
**Fax:** (613) 990-4901

## THESAURUS PROJECTS

### CANADIAN HERITAGE CORPORATE THESAURUS

**Status:** This project is at the information gathering stage.

**Comments:** Part of the vision for information management in the new department of Canadian Heritage is the creation of a corporate thesaurus. The Information Management Branch is still in the process of organising and staffing. No work has yet begun on the corporate thesaurus project.

**Contact:** Canadian Heritage
Glen Furgeson
**Phone:** (819) 991-0019
**Fax:** (819) 941-0977

### ENVIROSOURCE

**Comment:** The Envirosource project was established to build an information holdings architecture and access framework which would enable Environment Canada to provide easy and effective access to all its information. As part of this project, a number of thesaurus reports were developed. As well, controlled vocabularies in use at Environment were inventoried.

**Contact:** Environment Canada Library
**Phone:** (819) 997-1767
**Fax:** (819) 997-5349

### INFOSOURCE

**Major subject(s):** Subjects relevant to the Canadian federal government and general administration

**Language(s):** Bilingual: English and French

**Standard(s):** ISO

**Used by:** Public access will be made possible through public and private sector distributors

**Uses:** To manage and control a departmental inventory (government-wide) of information holdings; for document retrieval purposes of free/full text, bibliographic records and abstracts in an online database using the QL Systems Limited software

**Comments:** The InfoSource thesaurus is intended to function as a generic tool for sources of federal government information. It must be capable of being transported to other government systems.

**Contact:** Treasury Board Secretariat
Ross Hodgins
**Phone:** (613) 957-2486
**Fax:** (613) 957-8020

## MINISIS THESAURUS MANAGEMENT SYSTEM

**Comments:** The MINISIS thesaurus management system will consist of a core program supporting common thesaurus management functions. It will support the management of multilingual and monolingual thesauri, and the use of non-roman languages and alternate character sets, such as Arabic, Chinese and Japanese. The system will be able to interact with users in English, French, Spanish or other roman or non-roman languages. MINISIS will run on IBM-PC-compatible microcomputers, as well as Hewlett Packard/3000 minicomputers.

**Contact:** International Development Research Centre
Terry Gavin
**Phone:** (613) 236-6163, extension 2064
**Fax:** (613) 563-3858

## STANDARD DEPARTMENTAL SUBJECTS PROJECT

Standard subject list and data model.

**Uses:** Possibilities to use a thesaurus as an indexing and retrieval tool in free/full text format are being studied for use in 1995. An SGML tagger would also be used in the indexing process. The thesaurus will help manage and control the departmental inventory of information holdings.

**Comments:** This project is to identify a standard set of subjects for the department and to look at thesauri as tool in information location. Funding has been cut.

**Contact:** Industry Canada
Blair Stannard
**Phone:** (613) 952-2568
**Fax:** (613) 954-1894

**TERMINI**

**Major subject(s):** Diplomacy, legal language, international trade, and internal relations
**Language(s):** Bilingual: English and French

**Developed by:** Foreign Affairs and International Trade (FAIT)
**Date created:** Development undertaken in 1994
**Derived from:** TERMIUM

**Maintained by:** Foreign Affairs and International Trade

**Format(s):** ASCII
**Software/hardware used:** Basis/Basisplus
**Size:** 50,000

**Used by:** Users of the Corporate Automated Text Storage system
**Number of users:** 3,000
**Uses:** For indexing and retrieval purposes of free/full text in an online database using the Basis/Basisplus software

**Comments:** This is the thesaurus that we call our "uncontrolled" vocabulary. It includes terminology from TERMIUM plus additional terms from our department's documentation. Non-FAIT clients must approach TERMIUM management for permission to use.

**Contact:** Foreign Affairs and International Trade
Diane Crouse
**Phone:** (613) 996-9385
**Fax:** (613) 943-1090

## VERTICAL FILE THESAURUS

**Major subject(s):** Criminal justice
**Language(s):** Bilingual: English and French

**Developed by:** Library and Reference Centre of the Solicitor General Secretariat (SGS)
**Date created:** In the process of being developed since 1991
**Derived from:** National Criminal Justice Reference Service Thesaurus (NCJRS)
**Maintained by:** Solicitor General

**Format(s):** WordPerfect and paper
**Standard(s):** ISO standards in addition to the NCJRS standards

**Uses:** To index bibliographic records in an online database using the Procite software

**Comments:** This is a lightly adapted version of the NCJRS, enriched by French translation and some LC and local terms.

**Available from:** Solicitor General Secretariat
**Contact:** Heather Moore
**Phone:** (613) 991-2779
**Fax:** (613) 941-6171

## PRODUCT NAME INDEX

(i.e. name of thesauri, controlled vocabularies, thesaurus projects)

Aging Keywords

Aquatic Sciences and Fisheries Information System (ASFIS) Thesaurus

Archival Terminology

ARCHTHE

Bank of Canada List of Subject Headings

Biograph

Canadian Geographical Names Database (CGNDB)

Canadian Heritage Corporate Thesaurus

Canadian Inventory of Historic Buildings (CIHB)

Canadian Subject Headings (CSH)

Canadiana Authorities

CANDITH

CAN/MARC Country Codes

CAN/MARC Geographic Area Codes

CAN/MARC Language Codes

CATHES

Child Care Thesaurus

Coal Thesaurus

Conferences Authority List

CPIC (Canadian Police Information Centre)

Depository Services Program List of Subject Descriptors

Descriptive Cataloguing Standards for Philatelic Material: Thesaurus

DIS

Disability Related Subject Headings Thesaurus

Energy Thesaurus

Envirosource

Family Violence Thesaurus

Geoscan Thesaurus

Information Files/Periodical Index Collectives (Heading) Index

Information Files/Periodical Index Subject (Headings) Index and Thesaurus

Information Files/Periodical Index Individuals (Heading) Index

Information Services Glossary

Infosource

Integrated Staffing System Abbreviation List/Standard Terms for Naming Elements and System Component

Labourline Thesaurus

Library of Congress Subject Headings (LCSH)

Macrothesaurus for Information Processing in the Field of Economic and and Social Development (IDRC Version)

Management Keyword List

Manuscript Division Name Authority

Manuscript Division Subject Heading List

MINISIS Thesaurus Management System

Minproc Thesaurus

Mintec Thesaurus

Names Authority List

Naval Construction, 16th - 18th Centuries

Organisations Authority List

Park Nomenclature and Parks Visual Dictionary

Phocus Thesaurus

Posts Authority List

PRECIS

Répertoire des domaines TERMIUM

Standard Departmental Subjects Project

Statistics Canada Thesaurus

Subject Heading List for Bibliographies Database

Subject Heading List for Foreign Affairs Statements and Speeches/News Releases Database

TERMINI

Thésaurus des sciences et des technologies de l'information

Thesaurus Envirodoq

Thesaurus in Library and Information Science

Thesaurus of Subject Descriptors

THESMAS

Toponymie et terminologie en usage à Parcs Canada

Treaties Authority List

Trilingual Species Name List for the Waves/Vagues Database

Vertical File Thesaurus

Water Resources Thesaurus

# CORPORATE NAME INDEX

Bank of Canada

Canada Communication Group - Publishing

Canadian Centre for Management Development

Canadian Clearinghouse on Disability Issues

Canadian Heritage

Centre d'innovation en technologie de l'information

Environment Canada

Finance

Fisheries and Oceans

Foreign Affairs and International Trade

Geological Survey of Canada

Health Canada

Human Resources Development Canada

Industrie Canada

Industry Canada

International Development Research Centre

National Archives of Canada

National Defence

National Library of Canada

Natural Resources Canada

Parcs Canada

Parks Canada

Patrimoine canadien

Public Service Commission

Public Works and Government Services Canada

Royal Canadian Mounted Police

Solicitor General Secretariat

Statistics Canada

Travaux publics et services gouvernementaux Canada

Treasury Board Secretariat

## TSWG Sub-group III Member List

Diane Crouse
Foreign Affairs and International Trade

Christiane Desautels
National Archives of Canada

Jennifer Farkas
Canadian Information Technology Innovation
Industry Canada

Kerridwen Harvey
Canadian Heritage Information Network

Tony Moren
Statistics Canada

Rosemary Murray-Lachapelle (Chair)
National Archives of Canada

Sara Sprague
Human Resources Development

Judy Patterson
Canada Communication Group

Barry Sims
Government Telecommunications and Informatics Services

## Contact List

Diane Beattie
National Archives of Canada

Clément Bédard
National Parks Directorate

Jean Bennett
Environment Canada

Bruce Blair
Geological Survey of Canada

Lise Boudreault
Public Works and Government
Services Canada, TERMIUM

Susan Buggey
National Historic Sites

Ed Buchinski
Treasury Board Secretariat

Heather Cameron
Fisheries and Oceans

Colin Campbell
Foreign Affairs and International Trade

Céline Champagne
Status of Women Canada

Diane Crouse
Foreign Affairs and International Trade

Suzanne Cyr
National Archives of Canada

Mary Joan Dunn
Canadian Intelligence Security
Informatics Service

Leslie Firth
Bank of Canada

Glen Furgeson
Canadian Heritage

Terry Gavin
International Development Research Centre

Bruno Gnassi
Supply and Services Canada

Leslie Hamel
Labour Canada

Richard Harkin
Public Service Staff Relations Board

Rick Harrington
Public Service Commission

Fay Hjartarson
Statistics Canada

Ross Hodgins
Treasury Board of Canada

Frank Ianni
Human Resources Development

Sheila Burvill
Canadian Institute for Scientific and
Technical Information

Shayna Keces
Canadian Clearinghouse on Disability Issues

Helen Kerfoot
Canadian Permanent Committee on
Geographical Names

Pierre Landry
National Gallery of Canada

Roger Lepage
Foreign Affairs, Head of Cataloguing

Marty Lovelock
Health and Welfare Canada

Bill MacLennan
Agriculture Canada

Lorraine McQueen
Canadian Centre for Management

Diane Ménard
National Archives of Canada

Katherine Miller
Immigration and Refugee Board

Heather Moore
Solicitor General Canada

Marc Nason
Energy, Mines and Resources

Serge Paradis
Secretary of State

Velma Parker
National Archives of Canada

Frank Pope
Statistics Canada

Renita Repsys
National Defence

Nicole Sansfaçon
Canadian International Development Agency

Wendy Saunders
Environment Canada

Mary Short
Public Works and Government
Services Canada

Marcel Simoneau
Industry Canada
Canadian Information Technology for
Innovation

Doug Simpson
Royal Canadian Mounted Police

Maureen Sly
International Development Research Centre

Blair Stannard
Industry Canada

Hugo Stibbe
National Archives of Canada

Gerald Stone
National Archives of Canada

Ray Taylor
Patent Office

Jane Thacker
National Library of Canada

Lucie Thivierge
Canadian International Development Agengy

Bridget Turner
Fisheries and Oceans

Monique Viens
Canadian Advisory Council on the Status of Women

Pat Young
Canadian Heritage Information Network

Terri Tomchyshn
Status of Disabled Persons

## Cover Letter and Questionnaire

Date

Contact name

Dear ... :

Thank you for agreeing to complete the enclosed questionnaire. As discussed during our recent telephone conversation, we, on behalf of the Thesaurus Standards Working Group, are seeking your participation to help in the production of an inventory of thesauri, thesaurus projects and controlled vocabulary.

The Thesaurus Standards Working Group (TSWG) was established last October by the National Archives of Canada in collaboration with the Treasury Board Information Technology Standards Program. The purpose of the Working Group is to explore the use of thesauri as tools to help manage government information and inventories of information holdings. Based on its terms of reference and workplan, the TSWG instituted a sub-group whose mandate is to identify thesauri, thesaurus projects and controlled vocabulary in the Canadian federal government.

The inventory is intended to benefit those organisations eager to develop tools to enhance their ability to describe and preserve records in context, and to retrieve them in a cost-effective manner. Other objectives of the inventory are to share information, to draw upon the experience of others and to identify common sources and structures in order to avoid duplication of effort across the federal government. Ultimately, it is hoped that the information drawn from this questionnaire will assist departments and agencies in managing their information, in building inventories of information holdings as well as common systems.

The following explanations are intended to clarify the terms and objectives of the questionnaire.

- A **thesaurus** is a compilation of words and phrases showing synonymous, hierarchical and other relationships and dependencies, the function of which is to provide standardized vocabulary for information storage and retrieval (ANSI Z39.19 - 1980). In this questionnaire we seek information about thesauri, thesaurus projects and **controlled vocabulary** such as subject headings, geographic names and functions, consolidated indexes, and standard data values such as city codes and language codes which are used by departments and agencies to manage their own holdings.

- Many departments and agencies subscribe to commercial retrieval services which use thesauri in indexing and retrieval applications. Please note that this questionnaire is not intended to inventory these commercial retrieval services unless such a service is used to index the institutions' own information holdings.

Your response to this questionnaire is key to building a comprehensive inventory. Your contribution will help to ensure the success of the project. Completion of the questionnaire will take approximately fifteen minutes. Questions may be addressed to Christiane Desautels at (613) 947-1512. Thank you for your help.

Yours sincerely,

Rosemary Murray-Lachapelle
Chairperson, Thesaurus Standards Working Group
National Archives of Canada

Enclosure

Date

Personne-ressource

M/M...,

Au nom du Groupe de travail sur la normalisation des thésaurus (GTNT), nous sollicitons votre participation à l'établissement du répertoire des thésaurus, des projets de thésaurus et de vocabulaires contrôlés, comme mentionné lors de notre conversation téléphonique. Nous tenons donc à vous remercier d'avoir accepté de remplir le questionnaire en annexe.

En octobre dernier, les Archives nationales du Canada ont mis sur pied le GNTN, en collaboration avec le Programme du Conseil du Trésor portant sur les normes des technologies de l'information. Le but premier de ce Groupe vise à explorer l'utilisation du thésaurus en tant qu'outil de gestion de l'information gouvernementale et de répertoires de fonds de renseignements détenus par le gouvernement. Conformément à son mandat et à son plan de travail, le Groupe a mis sur pied un sous-groupe qui a pour tâche de dresser un inventaire des divers thésaurus, des projets de thésaurus et des vocabulaires contrôlés au sein du Gouvernement fédéral.

Cet inventaire a pour objet d'aider les institutions désireuses de mettre au point des outils de travail afin d'améliorer leur capacité à décrire et à conserver leurs documents avec toute information contextuelle pertinente, et à les repérer de manière rentable. Cet inventaire vise également l'échange d'informations, la mise en commun des expériences ainsi que le repérage des sources et des structures communes afin d'éviter la duplication des efforts au sein du Gouvernement. Nous aimerions que les résultats de ce travail de recherche puissent aider éventuellement les ministères et les organismes à mieux gérer l'information, à constituer des répertoires de leurs fonds de renseignements et à bâtir des systèmes communs.

Les explications qui suivent permettront, nous l'espérons, de préciser les objectifs du questionnaire et de vous familiariser avec la terminologie employée.

- Un **thésaurus** constitue une compilation de mots et de phrases ayant entre eux des relations de types synonymique, hiérarchique ou de dépendance ou tout autre type de relation, et qui a pour fonction de fournir un vocabulaire normalisé à des fins de stockage et de repérage de l'information.[1] Dans ce questionnaire, nous cherchons à obtenir des informations sur les thésaurus, les projets de thésaurus et les **vocabulaires contrôlés** tels les vedettes-matières, les noms et les fonctions géographiques, les index refondus, les données normalisées tels les codes de villes et de langues employés par les ministères et les organismes à des fins de gestion de leur propre fonds de renseignements.

- Plusieurs ministères et organismes abonnés à des services commerciaux pour le repérage de l'information utilisent des thésaurus pour l'indexation et le repérage de données. Veuillez noter que ce questionnaire ne cherche pas à répertorier ces services commerciaux à moins qu'un tel service ne serve à indexer les documents propres aux ministères et organismes.

Votre réponse à ce questionnaire constitue une contribution de première importance pour l'établissement d'un inventaire exhaustif. Votre appui contribuera à faire un succès de ce projet. Il ne vous faudra pas plus de quinze minutes pour y répondre. Pour toute question, veuillez communiquer avec Christiane Desautels au (613) 947-1512. Merci de votre collaboration!

Veuillez agréer l'expression de mes salutations distinguées.

La présidente du Groupe de travail sur la normalisation de thésaurus

Rosemary Murray-Lachapelle

Pièce jointe

---

[1] Traduction libre d'une définition d'un thésaurus tirée de la norme ANSI z39.19, 1980.

## Thesauri/Controlled Vocabulary Used in the Government of Canada As Tools to Manage Information

**Department/Agency** ______________________________

**Contact Name** ____________________ **Phone Number** ____________

**Job Title** ____________________ **Fax Number** ____________

- *Please answer questions with appropriate text, with a ✓ where applicable, or with NA if not applicable. For every section, please check as many options as are applicable.*
- *Additional information may be provided on a separate sheet. If you wish to report on more than one, please photocopy this form.*

1. **Type: Thesaurus(TH)**___ **Thesaurus Project(TP)**___ **Controlled Vocabulary(CV)**

   **Title**(or name) ______________________________

   ______________________________

   **If a CV, please specify: i) authority file** ___ **ii) code list** ___ **iii) keyword index** ___

   **iv) lexicon/dictionary** ___ **v) subject heading list** ___ **vi) other** ____________

   **Date of creation** ____________ **Date of latest revision** ____________

   **Updating frequency: monthly** __ **semi-annually** __ **annually** __ **as required** __

   **Developed by** (full corporate name) ______________________________

   **Maintained by** (full corporate name) ______________________________

   **Has it been derived from another TH/CV? Y**__ **N**__ **If so,**

   **which one?** ____________

   **TH/CV software/hardware** (if any) ______________________________

   **Available from:** ______________________________

   **Available in the following formats: i) paper** ___ **ii) microform** ___ **iii) CD ROM** ___

   **iv) other** (please specify coding format, e.g. ASCII, WordPerfect) ____________

2. **Major subjects** ______________________________

   **Minor subjects** ______________________________

   ______________________________

   **Number of terms** ______________________________

   **English** ___ **French** ___ **E/F bilingual** ___ **Other** ____________

   **Does it follow the ISO**(International Organisation for Standardization) **standards? Y**__ **N** __

   **Other standards? please specify** ______________________________

   **Used by** ____________________ **Number of users** _______

**Please return by fax to Christiane Desautels at (613) 947-1500. Thank you!**

## Thesauri/Controlled Vocabulary Used in the Government of Canada As Tools to Manage Information

**Questionnaire (cont'd)**

3. **The TH/CV is used:**

**a) As an indexing tool ___ If so, it is used with data in the following formats:**

**i) free/full text ___ ii) data in fields ___**

**iii) bibliographic records ___ iv) abstracts ___**

**v) other, please specify ______________________________**

**b) For document retrieval purposes ___ If so, it is used against the following formats:**

**i) free/full text ___ ii) data in fields ___**

**iii) bibliographic records ___ iv) abstracts ___**

**v) other, please specify ______________________________**

**c) In an online database ___ name of database software ____________**

**d) To help manage and control a departmental inventory of information holdings ___**

**e) Other, please specify ______________________________**

4. **General Comments ______________________________**

______________________________

______________________________

______________________________

______________________________

5. **Can this information be included in a thesaurus and controlled vocabulary inventory of the Government of Canada?**

**Y__ N__ DATE ____________ SIGNATURE ____________________**

6. **Can you help identify other TH/TP/CV in the federal government? If so, which ones? (Please provide contact names and telephone numbers, if available.)**

______________________________

______________________________

***Please return the questionnaire by fax to Christiane Desautels, National Archives of Canada, (613) 947-1500.***

***Thank you!***

## Thésaurus/Vocabulaire contrôlé en usage au Gouvernement fédéral comme outil de gestion de l'information

Ministère/Organisme ____________________

Personne-ressource ____________ Téléphone ____________

Titre de poste ____________ Télécopieur ____________

- Veuillez répondre aux questions en fournissant les renseignements pertinents, en cochant (✓) aux endroits appropriés ou en inscrivant SO, le cas échéant. Veuillez cocher toutes les réponses pertinentes pour chacune des sections.
- Si vous avez d'autres informations à nous fournir, veuillez les consigner sur une feuille séparée. S'il y a plus d'un projet à décrire, nous vous saurions gré de photocopier ce questionnaire en autant d'exemplaires requis.

1. Type: Thésaurus (TH)___ Projet de thésaurus (PT)___ Vocabulaire contrôlé (VC)___

Titre (ou nom) ____________________

____________________

Pour un VC, veuillez préciser: i) fichier d'autorité___ ii) liste de codes___

iii) index des mots-clés ___ iv) lexique/dictionnaire ___

v) liste de vedettes-matières ___ vi) autres ____________

Date de création ____________ Date de la dernière révision ____________

Fréquence de la mise à jour : mensuelle___ semestrielle___ annuelle___ au besoin___

Conçu par (denomination sociale au complet) ____________

Mise à jour effectuée par (denomination sociale au complet) ____________

Dérive-t-il d'un autre TH/VC? O___ N___ Le cas échéant, lequel? ____________

TH/VC logiciel/matériel (s'il y a lieu) ____________

Disponible auprès de: ____________

Disponible sur support: i) papier ___ ii) microforme ___ iii) CD ROM ___

iv) autres (préciser le format de codage, ASCII, WordPerfect, etc.) ____________

2. Sujets principaux ____________

Sujets secondaires ____________

____________________

Nombre de termes ____________

Anglais ___ Français___ Bilingue A/F ___ Autre ____________

Est-il conforme aux normes ISO(Organisation internationale de normalisation)? O___ N___

Autres normes? veuillez préciser ____________

Utilisé par ____________ Nombre d'utilisateurs ______

Veuillez retourner par télécopieur à Christiane Desautels au (613) 947-1500. Merci!

## Thésaurus/Vocabulaire contrôlé en usage au Gouvernement fédéral comme outil de gestion de l'information

**Questionnaire (suite)**

3. Le TH/VC est utilisé:

a) Comme outil d'indexation ___ si oui, selon quel format de présentation?
   i) texte intégral ___ ii) données dans des champs ___
   iii) fiches bibliographiques ___ iv) résumés analytiques ___
   v) autres, veuillez préciser ______________________________
b) Pour le repérage de documents ___ si oui, selon quel format de présentation?
   i) texte intégral ___ ii) données dans des champs ___
   iii) fiches bibliographiques ___ iv) résumés analytiques ___
   v) autres, veuillez préciser ______________________________
c) Dans une base de données en-direct ___ nom du logiciel ______________
d) Pour la gestion du répertoire des fonds de renseignements du ministère ________
e) Autre, veuillez préciser ______________________________

4. Remarques ______________________________________________
______________________________________________
______________________________________________
______________________________________________
______________________________________________

5. Peut-on inclure ces informations dans un répertoire de thésaurus et de vocabulaires contrôlés du Gouvernement fédéral?
O___ N___ DATE ______________ SIGNATURE ______________________

6. Pourriez-vous nous aider à repérer d'autres TH/PT/VC au Gouvernement fédéral? Le cas échéant, lesquels? (Veuillez fournir le nom des personnes-ressources avec leurs numéros de téléphone, s'il y a lieu.)
______________________________________________
______________________________________________

Veuillez retourner ce questionnaire par télécopieur à Christiane Desautels
Archives nationales du Canada, (613) 947-1500.

Merci!

## 9. STANDARDS FOR THESAURUS, TERMINOLOGY, VOCABULARY, AND INDEXING

(Prepared by the National Archives of Canada)

1. **Standards For Monolingual Thesaurus**
2. **Standards For Multilingual Thesaurus**
3. **Standards For Terminology and Vocabulary**
4. **Standars For Indexing**

## 1. Standards For Monolingual Thesaurus

1.1 **International Organization for Standardization. *ISO Standards Handbook 1 - Documentation and Information.* 3rd ed. Geneva: ISO, 1988.**

**Organisation internationale de normalisation. *Recueil de normes ISO 1 - Documentation et information.* 3e éd. Genève: ISO, 1988.**

This handbook contains the complete texts of ISO standards concerning information collection, processing, storage and retrieval. It is divided into six parts: vocabularies; terminology (principles and coordination); character sets and transliteration; library science and documentation; documents in administration, commerce and industry; and documentary reproduction. Standards for electronic data processing are contained in *ISO Standards Handbook 9 - Data Processing - Software*.

1.2 **International Organization for Standardization. *Documentation - Guidelines for the Establishment and Development of Monolingual Thesauri.* 2nd ed. Geneva: ISO, 1986. Ref. no.: ISO 2788:1986 (E)**

**Organisation internationale de normalisation. *Documentation - Principes directeurs pour l'établissement et le développement de thésaurus monolingues.* 2e éd. Genève: ISO, 1986. Réf. no.: ISO 2788:1986 (F).**

This ISO standard is a basic, clearly written guide to the construction of a monolingual thesaurus and includes a basic explanation of thesaurus, definitions, abbreviations, symbols, types of relationships, types of thesaurus displays, methods of compilation, construction, maintenance, and management.

1.3 **Association française de normalisation. *Documentation - Règles d'établissement des thésaurus en langue française.* 2$^e$ éd. Paris: AFNOR, 1981. (AFNOR Z 47-100: décembre 1981).**

This French standard is in accordance with ISO 2788 although slight modifications have been made to the ISO standard in order to apply it to the construction of a French-language thesaurus. A very helpful flowchart on how to build a thesaurus summarizes, in one page, the process of thesaurus construction.

1.4 **British Standards Institution. *British Standard Guide to Establishment and Development of Monolingual Thesauri.* 2nd ed. London: BSI, 1987. (BS 5723:1987).**

This standard is identical to ISO 2788.

1.5 **United Nations Educational, Scientific and Cultural Organization.** ***UNISIST Guidelines for the Establishment and Development of Monolingual Thesauri.*** **2nd ed. Paris: PGI, September 1981. (PGI-81/WS/15).**

This document was prepared as a basis for ISO 2788, and is superseded by the ISO standard.

1.6 **American National Standards Institute.** ***Guidelines for Thesaurus Structure, Construction and Use.*** **New York: ANSO, 1980. (ANSI/NISO Z39.19-1980).**

This standard provides guidelines for constructing monolingual thesauri and is consistent with ISO 2788.[2]

## 2. Standards For Multilingual Thesaurus

2.1 **International Organization for Standardization.** ***Documentation - Guidelines for the Establishment and Development of Multilingual Thesauri.*** **1st ed. Geneva: ISO, 1985. Ref. no.: ISO 5964:1985 (E).**

**Organisation internationale de normalisation.** ***Documentation - Principes directeurs pour l'établissement et le développement de thésaurus multilingues.*** **1ère éd. Genève: ISO, 1985. Réf. no.: ISO 5964: 1985 (F).**

This document concentrates on the development, display and management of multilingual thesauri where the requirements differ from those of monolingual thesauri. Examples and displays are presented in English, French and German. This standard should be used in conjunction with ISO 2788.

2.2 **Association française de normalisation.** ***Documentation - Principes directeurs pour l'établissement des thésaurus multilingues.*** **1ère éd. Paris: AFNOR, 1990. (AFNOR Z 47-101: décembre, 1990).**

This French standard is identical to ISO 5964 except for minor editorial language modifications. This document should be used in conjunction with AFNOR Z 47-100.

---

[2] This standard was rewritten and circulated for approval in 1990 as ANSI/NISO Z39-19-199x American National Standard Guidelines for Thesaurus Construction, Structure and Use, with the following abstract: "This standard provides guidelines for constructing monolingual thesauri: formulating the descriptors, displaying relationships among terms, and effectively presenting the information to the user in print and online. Recommendations for capabilities of thesaurus software are also included." It was withdrawn 6 June 1993 and a publication date is not yet available.

2.3 **British Standards Institution.** ***British Standard Guide to Establishment and Development of Multilingual Thesauri.*** **1st ed. London: BSI, 1985. (BS 6723:1985).**

This standard is identical to ISO 5964.

2.4 **United Nations Educational, Scientific and Cultural Organization.** ***UNISIST Guidelines for the Establishment and Development of Multilingual Thesauri.*** **Paris: PGI, September 1981. (PGI-81/WS/12).**

This standard is the forerunner of ISO 5964 and extends the guidelines set out in ISO 2788.

## 3. Standards For Terminology and Vocabulary

3.1 **International Organization for Standardization.** ***Documentation and Information - Vocabulary - Part 6: Documentary Languages.*** **1st ed., bilingual. Geneva: ISO, 1983. Ref. no.: ISO 5127-6:1983 (E/F).**

**Organisation internationale de normalisation.** ***Documentation et information - Vocabulaire -partie 6: Langages documentaires.*** **1ère éd., bilingue. Genève: ISO, 1983. Réf. no: ISO 5127-6:1983 (E/F).**

This part of the standard describes the basic concepts for information and documentation, such as notations, classifications and thesauri, which are used for describing documentary languages.

3.2 **Association française de normalisation.** ***Documentation - Thésaurus monolingues et multilingues - symbolisation des rélations.*** **1ère éd. Paris: AFNOR, 1980. (AFNOR Z 47-103: avril 1980).**

This standard describes the symbols to be used in the various parts of monolingual and multilingual thesauri to represent the conventional relationships between terms. The symbols are applicable to thesauri in paper, in microform or in electronic form and are compatible with ISO 1951.

3.3 **British Standards Institution.** ***Glossary of Documentation Terms.*** **London: BSI, 1976. (BS 5408:1976).**

This standard was adopted as technically equivalent to the ISO 5127/6:1983.

3.4 **International Organization for Standardization. *Terminology - Vocabulary*. 1st ed., bilingual. Geneva: ISO, 1990. Ref. no.: ISO 1087:1990) (E/F).**

**Organisation internationale de normalisation. *Terminologie - Vocabulaire*. 1ère éd., bilingue. Genève: ISO, 1990. Réf. no: ISO 1087:1990 (E/F).**

This standard contains the basic terms used in terminology and lexicography and is recommended for all terminological work in the field of standardization. The majority of the terms correspond to ISO R704.

3.5 **International Organization for Standardization. *Principles and Methods of Terminology*. 1st ed. Geneva: ISO, 1987. Ref. no.: ISO 704:1987 (E).**

**Organisation internationale de normalisation. *Principes et méthodes de la terminologie*. 1ère éd. Genève: ISO, 1987. Réf. no: ISO 704:1987 (F).**

This standard provides the principles required to unify, standardize or create new concepts and terms in the scientific and technological fields, and is to be used in conjunction with ISO 1087.

3.6 **Canadian Standards Association. *Information Technology Vocabulary*. Toronto: CAN/CSA, 1992. CAN/CSA-Z243-58-92). (E/F)**

This Canadian national standard is a bilingual dictionary and contains basic vocabulary describing fundamental computer systems technology with terms related to such technological developments as OSI and expert systems.

3.7 **Canadian Standards Association. *Principles and Methods of Terminology*. Toronto: CAN/CSA, 1992. CAN/CSA-Z780-92). (E/F)**

This Canadian national standard was approved by the Standards Council of Canada and is an adoption of ISO 704.

3.8 **Canadian Standards Association. *Terminology - Vocabulary*. Toronto: CAN/CSA, 1992. (CAN/CSA-Z781-92). (E/F)**

This Canadian national standard was approved by the Standards Council of Canada and is an adoption of ISO 1087.

3.9 **International Organization for Standardization. *Information Processing Systems - Vocabulary - Part 06: Preparation and Handling of Data*. Geneva: ISO, 1987. Ref.: no.: ISO 2382-6:1987 (E/F).**

**Organisation internationale de normalisation.** ***Systèmes de traitement de l'inforamtion - Vocabulaire - Partie 06: Préparation et manipulation des données.*** **2e éd., bilingue. Genève: ISO, 1987. Réf. no: ISO 2382-6:1987 (E/F).**

This standard deals with input and output of data, transfer and conversion methods, and search techniques by presenting, in English and French, terms and definitions of selected concepts relevant to the field of information processing and identifies the relationships between entries. It is intended to facilitate international communication in information processing in order to clarify concepts, select terms to be used in various languages or in various countries to express the same concept, and establish definitions to provide satisfactory equivalents for the various terms in different languages.

## 4. Standards For Indexing

4.1 **International Organization for Standardization.** ***Documentation - Methods for Examining Documents, Determining Their Subjects, and Selecting Indexing Terms.*** **1st ed. Geneva: ISO, 1985. Ref.: no.: ISO 5963:1985 (E).**

**Organisation internationale de normalisation.** ***Documentation - Méthodes pour l'analyse des documents, la détermination de leur contenu et la sélection des termes d'indexation.*** **1ère éd. Genève: ISO, 1985. Réf. no: ISO 5963:1985 (F).**

This Standard is intended primarily as a guide to indexers, during the stages of document analysis and concept identification, and describes recommended procedures for examining documents, determining their subjects, and selecting appropriate indexing terms. It may also be helpful for the analysis of users' enquiries and their translation, for retrieval purposes, into the controlled terms of an indexing language.

4.2 **British Standards Institution.** ***British Standard Recommendations for Examining Documents, Determining Their Subjects and Selecting Indexing Terms.*** **2nd ed. London: BSI, 1991. (BS 6529:1991).**

This standard has been adopted as equivalent to ISO 5963:1985.

4.3 **American National Standards Institute.** ***American National Standard for Library and Information Sciences and Related Publishing Practices - Basic Criteria for Indexes.*** **New York: ANSO, 1984. (ANSO Z39.4-1984)**

This standard is consistent with ISO 5963:1985.

4.4 **International Organization for Standardization.** ***Information and documentation - Guidelines for the content, organization and presentation of indexes*****. 2nd ed. Geneva: ISO, 199x. Ref.: no.: ISO 999 (E).**

**Organisation internationale de normalisation.** ***Information et documentation - Principes directeurs pour l'élaboration, la structure et la présentation des index*****. 2e éd. Genève: ISO, 199x. Réf. no: ISO 999 (F).**

This standard replaces the 1975 edition of ISO 999. It covers basic indexing principles and practices, including the choice, form and arrangement of headings used in index entries.

## 10. THESAURUS: AN ANNOTATED BIBLIOGRAPHY

(Prepared by the National Archives of Canada)

1. **Thesaurus General**
2. **Thesaurus Construction and Maintenance**
3. **Thesaurus and Other Retrieval Techniques**
4. **Software For Thesaurus**

INDEX

## 1. Thesaurus General

1.1 **Aitchison, Jean and Gilchrist, Alan. *Thesaurus Construction: A Practical Manual.* 2d ed. London: The Association for Information Management, 1987.**

This is a practical and concise guide to the construction of thesauri for use in information retrieval, from planning and design to maintenance and updating.

1.2 **Chen, H., Lynch, K.J., Basu, K., Ng, T.D. Generating, Integrating and Activating Thesauri for Concepts-Based Document Retrieval, *IEEE Expert.* (1993): 25-33**

This paper proposes an approach that uses an algorithm based on heuristics that can activate related terms across multiple thesauri.

1.3 **Fugman, Robert. "An Interactive Classaurus on the PC." *International Classification* 17, no. 3/4 (1990): 133-37.**

The construction of an index or indexing language to retrieve information consists of a vocabulary which may be either a thesaurus or a classification system. The author discusses the requirments for the various conceptual categories and suggests the integration of natural language terms with a controlled vocabulary to form a hybrid - a classaurus.

1.4 **Grefenstette, G. *Explorations in Automatic Thesaurus Discovery.* Norwell, MA: Kluwer Academic Publishers, 1994.**

This book proposes innovative techniques relevant to large-scale information systems and methods for identifying classes of related lexical items. It also features evaluation techniques to assess the results of semantic-relation discovery.

1.5 **Jones, Susan. "A Thesaurus Data Model for an Intelligent Retrieval System." *Journal of Information Science* 19, no. 3 (1993): 1-10.**

This article describes a project where two standard thesauri were loaded into a relational database management system to explore characteristics and to experiment with methods of user navigation as an aid to document retrieval. The data modelling and database design techniques used are described.

1.6 **Lancaster, Frederick William.** ***Vocabulary Control for Information Retrieval.*** **2d ed. Virginia: Information Resources Press, 1986.**

This is a good detailed primer for the understanding of thesauri: the various types, construction, composition, and uses.

1.7 **McLelland, Joe. "Computers, Databases and Thesauri."** ***Aslib Proceedings*** **42, no. 7/8 (July/August 1990): 201-5.**

Two different approaches for information retrieval and how to combine them are described: the view of an information specialist where the database management system offers a thesaurus module as one component of the package; and the view of the systems analyst/programmer where specialist thesaurus handling packages or services are available which operate without reference to a specific database management system.

1.8 **Meadows, Jack. "Purposes and practices of text retrieval."** ***Journal of Information Science*** **18, no. 2 (1992): 83-87.**

The author discusses different methods of retrieval, restrictions to retrieval and different solutions. These include retrieval through full text, use of graphics, use of equations, and SGML; the restrictions of end user requirments and the type of database; and solutions including the addition of expert systems and thesaurus generation.

1.9 **Pollitt, A. Steven.** ***Information Storage and Retrieval Systems: Origin, Development and Applications.*** **Chichester: Ellis Harwood Limited, 1989.**

This is a standard textbook on information retrieval covering such areas as classifying and indexing systems, searching manual and automated systems, building and searching a database and front-end systems.

1.10 **Schmitz-Esser, Winfried. "New Approaches in Thesaurus Application."** ***International Classification*** **18 no. 3 (1991): 143-47.**

This article describes new approaches for thesauri applications including expert and interface systems, object-oriented design and programming, hypertext systems, machine translation and machine abstracting.

## 2. Thesaurus Construction and Maintenance

2.1 **Bertrand-Gastaldy, Suzanne and Davidson, Colin H. "Improved Design of Graphic Displays in Thesauri - Through Technology and Ergonomics." *Journal of Documentation* 42, no. 4 (December 1986): 225-51.**

This article presents a list of performance criteria that an 'ideal' thesaurus graphic display should respect.

2.2 **Crawford, R.G. Automatic Thesaurus Construction Based on Term Centroids. *The Canadian Journal of Information Science*, vol. 4 (1979): pp. 124-136.**

This paper examines the problem of what properties of terms are to be used to group terms and the concepts of term centroids. It then presents the retrieval results obtained from a fully automatically constructed thesaurus.

2.3 **Crouch, C.J. An Approach to the Automatic Construction of Global Thesauri. *Information Processing & Management*, vol. 26, no. 5 (1990): pp. 629-640**

This paper reviews approaches to automatic thesaurus construction and, based on the term discrimination value model from Salton, Yang and Yu, presents an approach to the automatic generation of global thesaurus. This method, through its test application to two document collections, shows improvements in retrieval activities.

2.4 **Field, B.J. *Semi-automatic Development of Thesauri Using Free-language Vocabulary Analysis (Part 1 only) Report No. R75/24.* London: The Institution of Electrical Engineers, July 1975.**

This is a highly technical paper detailing the development of a structured vocabulary system or thesaurus, based on free language indexing. The development of an automatic truncation procedure is also described and includes the results of statistical studies on the growth and frequency characteristics of the free-language vocabulary.

2.5 **McMath, Charles F., Tamaru, Robert S., Rada, Roy. "A graphical thesaurus-based information retrieval system." *International Journal of Man-Machine Studies* 31 (1989): 121-47.**

This article describes the use of computer generated graphic displays using a thesaurus model for concept evaluation and term relationships and based on a thesaurus approach to information retrieval.

**2.6 Rada, Roy. "Maintaining Thesauri and Metathesauri."** ***International Classification*** **17, no. 3/4 (1990): 158-64.**

The easiest way to connect different databases is to map together the terms of their respective thesauri creating a metathesaurus. This article describes a number of systems which support thesauri and metathesauri and provides good examples for creating a metathesaurus. Particular references are made to biomedical systems and thesauri.

**2.7 Rees-Potter, Lorna Katherine. "Dynamic Thesaural Systems."** ***Canadian Library Journal*** **46, no. 1 (February 1989): 40-41.**

The author describes how terminology in citation contexts is useful for concept retrieval in specialty areas in the social sciences where the terminology is less precise. The author also indicates that further research is being carried out to determine the relationships among the speciality area terms in order to construct a concept record which will specify the term and its synonyms, its hierarchy, and any other relationship that might be of interest.

## 3. Thesaurus and Other Retrieval Techniques

**3.1 Basch, Reva. "The Seven Deadly Sins of Full-Text Searching."** ***Database*** **12, no. 4 (August 1989): 15-23.**

Seven problem areas associated with full text searching are described and include online versus offline searching, lack of controlled vocabulary, inadequate software, lack of relevance of displayed information, data entry errors, searching incompatible databases, and different online formats for the same information. Different types of software and systems are discussed under each problem area and references and recommendations for additional reading are included at the end of the article.

**3.2 Blair, David C. "An Extended Relational Document Retrieval Model."** ***Information Processing and Management*** **24 (1988): 349-71.**

This is a good technical primer for information/document retrieval based on an extended relational logical model. Retrieval considerations such as associative searching and inferential retrieval are discussed.

**3.3 Chen, Hsinchun. "Knowledge-based Document retrieval: Framework and Design."** ***Journal of Information Science*** **18 (1992): 293-314.**

This author explores a prototype knowledge-based retrieval system for intelligent information retrieval using various system-supported procedures and aids. It makes extensive use of the users' and the information specialists' knowledge and heuristics, it attempts to alleviate the user's misconceptions during the search, it adopts the strengths of the existing inverted index-based systems, and utilizesf a uniform and coherent blackboard architecture.

3.4 **Coco, Al. "Full Text Revisited."** ***Legal Reference Services Quarterly*** **6, no. 3-4 (June 1987): 195-97.**

This short article praises the addition of a controlled vocabulary to full text which improves retrieval quality due to the inclusion of extra terms, and the use of better editorial components in legal headnotes. Search terms can be combined in closer proximity improving the precision factor.

3.5 **Comber, Denis, and Stanford, Joy. "Free Text and Thesaurus Control".** ***The International Records Management Council Journal*** **(May 1993): 8-13.**

A very good article which discusses information retrieval using free text and using a thesaurus controlled vocabulary to search an online records management database. An annotated bibliography of related articles and texts is included.

3.6 **Keen, Michael E. "Some Aspects of Proximity Searching in Text Retrieval Systems."** ***Journal of Information Science*** **18, no 2 (1992): 89-98.**

The capabilities, syntax and circumstances for the use of proximity searching are discussed, as well as the relevancies of precision and recall and their impact on retrieval.

3.7 **Paice, Chris D. "A Thesaural Model of Information Retrieval." Information Processing & Management 27, no. 5 (January 1991).**

This paper suggests that topic representation be drawn upon excerpts from a domain thesaurus. It proposes an algorithm for doing so and examines if thesauri can serve this purpose.

3.8 **Perez, Ernest. "Text Enhancement-Controlled Vocabulary vs. Free Text."** ***Special Libraries*** **73, no. 3 (July 1982): 183-92.**

Full or free text online access to and retrieval from newspaper databases is advocated in this article. The possibilities of including limited controlled vocabulary lists is also discussed.

3.9 **Pritchard, John. "Electronic Filing and Retrieval: Developments in Full Text Retrieval Systems."** ***Information Media and Technology*** **23, no. 2 (March 1990): 63-66.**

This article is a good primer on full text retrieval: what it does and how it is done. It describes the differences between database management systems, full text retrieval, free text retrieval, and the applications which are best served by each form of retrieval. Also mentioned are optical character recognition systems (OCR), imaging systems and hypertext.

3.10 **Schuegraf, Ernest J. and Martin F. van Bommel. "An Automatic Document Indexing System Based on Cooperating Expert Systems: Design and Development."** ***The Canadian Journal of Information and Library Science*** **18, no. 2 (July/August 1993): 32-50.**

The subject system of this article involves a combination of a pre-processing statistical module identifying the content-bearing words and the expert system Topic Expert to determine the subject areas of the document from these words by matching them to vocabularies characteristic of given subject areas. Another expert system, the Subject Specialist, assigned final index terms according to the subject areas and the content-bearing words by using an enriched thesaurus.

## 4. Software For Thesaurus

4.1 **Chan, Lois Mai and Pollard, Richard.** ***Thesauri Used in Online Databases: An Analytical Guide.*** **Westport: Greenwood Press, 1988.**

Many individual databases have developed specific controlled vocabularies in their thesauri because of the diversity in subject coverage and purpose. The purpose of this guide is to provide assistance in the identification of indexing vocabularies used by large and/or widely used online databases.

4.2 **Chong, Alicia. "Topic: A Concept-based Document Retrieval System."** ***Library Software Review (USE)*** **8 (1989): 281-84.**

This article critiques the Verity, Inc. product, Topic which is a document management, storage and retrieval system based on concept retrieval.

4.3 **Ganzmann, Jochen. "Criteria for the Evaluation of Thesaurus Software."** ***International Classification*** **17, no. 3/4 (1990): 148-57.**

The author defines criteria for thesaurus software in order to help the end-user select software appropriate to specific needs. He explains criteria for the major functions of thesaurus construction, maintenance, and updating. Additional functionality requirements are also presented in an appendix, Check-list for Thesaurus Software.

4.4 **Milstead, Jessica L. "Thesaurus Software Packages for Personal Computers."** ***Database*** **13, no. 6 (December 1990): 61-65.**

The author discusses six thesaurus management packages available in the United States and provides a comparison of these packages. She also discusses integrating a thesaurus module with the entire corporate database management system as opposed to having the module operate only a part of the system.

4.5 **Ritzler, Claus. "Comparative Study of PC-supported Thesaurus Software."** ***International Classification*** **17, no. 3/4 (1990) 138-47.**

Three PC supported software packages are compared for the development, construction and management of thesauri and other word material with emphasis on hardware and software requirements, handling and user interface, functionality and reliability.

## INDEX
### (by author)

Aitchison, Jean and Gilchrist, Alan
Thesaurus Construction: A Practical Manual

Basch, Reva
The Seven Deadly Sins of Full-Text Searching

Bertrand-Gastaldy, Suzanne and Davidson, Colin H.
Improved Design of Graphic Displays in Thesauri

Blair, David C.
An Extended Relational Document Retrieval Model

Chan, Lois Mai and Pollard, Richard
Thesauri Used in Online Databases: An Analytical Guide

Chen, H., Lynch, K.J., Basu, K., Ng, T.D.
Generating, Integrating and Activating
Thesauri for Concepts-Based Document Retrieval

Chen, Hsinchun
Knowledge-based Document Retrieval: Framework and Design

Chong, Alicia
Topic: A Concept-based Document Retrieval System

Coco, Al
Full Text Revisited

Comber, Denis, and Stanford, Joy
Free Text and Thesaurus Control

Crawford, R.G.
Automatic Thesaurus Construction Based on Term

Crouch, C.J.
An Approach to the Automatic Construction of Global Thesauri

Field, B.J.
Semi-automatic Development of Thesauri ...

Fugman, Robert
An Interactive Classaurus on the PC

Ganzmann, Jochen
Criteria for the Evaluation of Thesaurus Software

Grefenstette, G.
Explorations in Automatic Thesaurus Discover

Jones, Susan
A Thesaurus Data Model for an Intelligent Retrieval System

Keen, Michael E.
Some Aspects of Proximity Searching in Text Retrieval Systems

Lancaster, Frederick William
Vocabulary Control for Information Retrieval

McLelland, Joe
Computers, Databases and Thesauri

McMath, Charles F., Tamaru, Robert S., Rada, Roy
A Graphical Thesaurus-based Information Retrieval System

Meadows, Jack
Purposes and Practices of Text Retrieval

Milstead, Jessica L.
Thesaurus Software Packages for Personal Computers

Perez, Ernest
Text Enhancement-Controlled Vocabulary vs. Free Text

Pollitt, A. Steven
Information Storage and Retrieval Systems: Origin, Development ...

Pritchard, John
Electronic Filing and Retrieval: Developments in Full Text ...

Rada, Roy
Maintaining Thesauri and Metathesauri

Rees-Potter, Lorna Katherine
    Dynamic Thesaural Systems

Ritzler, Claus
    Comparative Study of PC-supported Thesaurus Software

Schmitz-Esser, Winfried
    New Approaches in Thesaurus Application

Schuegraf, Ernest J. and Martin F. van Bommel
    An Automatic Document Indexing System Based on Cooperating …

# Appendix A

## TBITS Thesaurus Standards Working Group:

## Official Documents

## Thesaurus Standards Working Group
## Members 1993-1994

Edite Abols
Library, Justice Canada

Janet Beale
Government Services Canada

Claude-André Bonin
Hydro Québec

Lise Boudreault
Public Works and Government Services

Bob Cheetham
Health Canada

Diane Crouse
External Affairs and International Trade

Bob Derksen
Revenue Canada

Fay Hjartarson
Statistics Canada

Ross Hodgins
Treasury Board Secretariat

Lucie Joyal-Serhal
Consulting and Audit Canada

J.P.W. Larocque
Canadian Heritage Information Network

Micheline Latulippe
Hydro Québec

Paul McLaughlin
National Defence

Lucie Molgat
Labour Canada (HRLC)

Christiane Desautels
National Archives of Canada

Keldine FitzGerald
Human Resources Development

Glen Furgeson
Canadian Heritage Information Network

Ron Gauthier
Environment Canada

Sue Gavrel
Treasury Board Canada

Bruno Gnassi
Canada Communications Group

Kerridwen Harvey
Canadian Heritage Information Network

Rosemary Murray-Lachapelle (Chair)
National Archives of Canada

Maryna Nowosielski
Human Resources Development

John Nugent
Patent Office

Judy Patterson
Canada Communications Group

Wendy Saunders
Environment Canada

Barry Sims
Government Telecommunications Information Services

Maureen Sly
International Development Research Centre

Tony Moren
Environment Canada

Blair Stannard
Industry, Science Canada

Jane Thacker
National Library of Canada

Roy Vallillee
Natural Resources Canada

Pat Young
Canadian Heritage Information Network

## Members of Sub-group I:
## Definition of Requirements

Janet Beale
Government Services Canada

Glen Furgeson
Canadian Heritage Information Network

Ross Hodgins
Treasury Board Secretariat

Linda Morrisey
Human Resources Development

Blair Stannard
Industry, Science Canada

Christiane Desautels
National Archives of Canada

Sue Gavrel
Treasury Board Secretariat

Candace Loewen
National Archives of Canada

Rosemary Murray-Lachapelle (Chair)
National Archives of Canada

## Members of Sub-group II: Standards Review

Rosemary Murray-Lachapelle
National Archives of Canada

Keldine FitzGerald
Human Resources Development

Beth Knox
National Archives of Canada

Maureen Sly
International Development Research Centre

Edite Abols
Justice Canada

Jane Thacker
National Library of Canada

Lise Boudreault
Public Works and Government Services

Fay Hjartarson
Statistics Canada

## Members of Sub-group III:

Diane Crouse
Foreign Affairs and International Trade

Jennifer Farkas
Canadian Information Technology Innovation
Industry Canada

Tony Moren
Statistics Canada

Sara Sprague
Human Resources Development

Barry Sims
Government Telecommunications and
Informatics Services

Christiane Desautels
National Archives of Canada

Kerridwen Harvey
Canadian Heritage Information Network

Rosemary Murray-Lachapelle (Chair)
National Archives of Canada

Judy Patterson
Canada Communications Group

## Thesaurus Standards Working Group

### Editorial Sub-group

Rosemary Murray-Lachapelle (Chair)
National Archives of Canada

Christiane Desautels
National Archives of Canada

Jane Thacker
National Library of Canada

Maureen Sly
International Development Research Centre

Pat Young
Canadian Heritage Information Network